Monitoring in language perception

An electrophysiological investigation

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op vrijdag 29 augustus 2008
des namiddags om 1.30 uur precies

door

Constance Theodora Wilhelmina Maria Vissers

geboren op 3 november 1979
te Bergen op Zoom

The studies described in this thesis were carried out at the Nijmegen Institute of Cognition and Information (NICI), Radboud University Nijmegen, The Netherlands.

ISBN 978-90-9023006-1

Copyright © by Constance Th.W.M. Vissers.

Design front cover: Wendy Bergervoet
Lay out: In Zicht Grafisch Ontwerp, Arnhem
Printed and bound by PrintPartners Ipskamp

Promotor

Prof. dr. H.H.J. Kolk

Co-promotor

Dr. D.J. Chwilla

Manuscriptcommissie

Prof. dr. H.J. Schriefers

Prof. dr. L. Osterhout (University of Washington)

Dr. L. A. Stowe (Rijksuniversiteit Groningen)

Voor mijn ouders

Paranimfen

Dr. W.H.P.M. Vissers

Mr. M. van der Meulen

Contents

Chapter 1

General Introduction

Chapter 1: General Introduction

To err is human. This familiar adage is as true in the area of language as in any other human behaviour. Therefore, it is important for the cognitive system to safeguard against the consequences of errors. Before a mistake can be corrected, it must be detected. In the domain of language production, it has been claimed that an online checking mechanism or ‘monitor’ exists to detect trouble and initiate self-repair. This process of watching over the quality of ones performance leads to output optimalization (e.g., Stuss & Benson, 1986). In this dissertation, I propose that a similar monitoring process also operates in the domain of language perception.

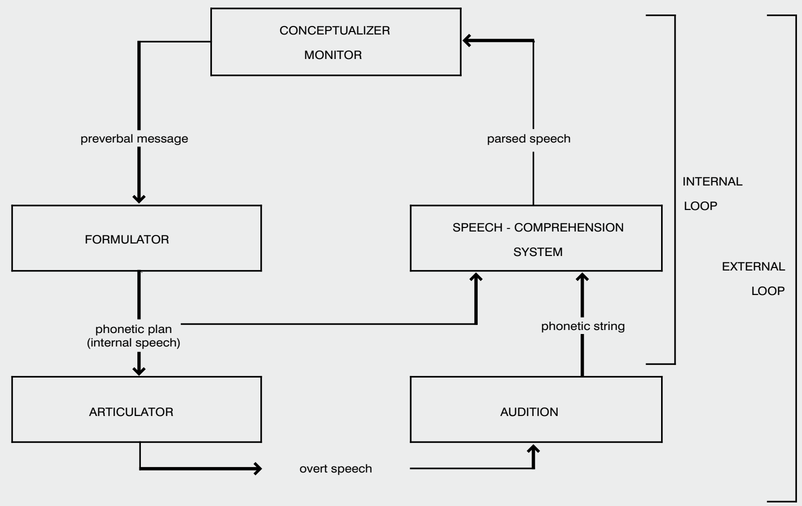
Monitoring in language production

The human ability to self-monitor can, for instance, manifest itself in speaking. Speaking involves a number of complex speech planning processing stages. The first step is to prepare the message. The speaker determines *what* he wants to express in his message to the listener and also *how* this information will be expressed. After that, the speaker formulates the conceptual message. This involves putting the message in a grammatically and phonologically correct form. In the last stage of speech production, a plan for how an utterance has to be articulated has to be made. The word’s syllables are mapped onto motor patterns that move the tongue, mouth and vocal apparatus in order to generate overt speech. Given the highly complex nature of these processing stages, much can go wrong in speaking. Still, errors are rare in ordinary speech. Garnham, Shillcock, Brown, Mill, and Cutler (1981) found a total number of 191 slips of the tongue in a text corpus of 200000 words; this is no more than about one speech error per thousands words. Despite this low error rate, we continuously keep control of what we are saying and how we are saying it. If a speaker makes a mistake, or expresses something inappropriately, he can spontaneously repair his own speech (Levelt, 1983). An important question is: How is this monitoring mechanism for spoken language organized?

Levelt’s (1989) theory of monitoring in language production proposes that a speaker has access to both his internal speech and his overt speech by means of the

so-called Speech-Comprehension System. As Levelt (1989, p.13) puts it, “a speaker is his own listener.” A speaker can thus listen to his own speech, just as he can listen to the speech of others. Levelt (1983) proposes a double ‘perceptual loop’ for the Speech-Comprehension System. Specifically, he proposes an internal loop involving perception of internal speech before it is uttered (the phonetic plan) and an external loop involving perception of self-produced overt speech. So, according to Levelt’s perceptual-loop theory, both internal speech and overt speech are fed into the Speech-Comprehension System for checking purposes. This view of self-monitoring is depicted in Figure 1.

Figure 1 Levelt’s perceptual loop theory of self-monitoring



Monitoring of overt speech can manifest itself with the phenomenon of ‘overt’ self- repairs, in which speech is interrupted and a new attempt is made at producing the correct form, as in the following example (Levelt & Cutler, 1983):

‘What things are this kid - is this kid going to say incorrectly?’

Here the speaker noticed an error of subject-verb agreement and corrected it. The monitoring of internal speech is evident in ‘covert’ repairs, in which errors are intercepted at the level of planning. In this way, the speaker can detect trouble in his own internal speech before articulation. Covert repairs become apparent as various speech dysfluencies such as hesitations, prolongations or pauses. Note the following example from Schegloff (1979):

'Tell me, uh what - d'you need a hot sauce?'

Here the speaker probably started out intending to say *do you need*?, but it was apparently more appropriate to issue a Yes/No question. The original utterance was therefore interrupted and a fresh start was made.

As described above, according to Levelt's perceptual loop theory, we use the same Speech-Comprehension System for comprehending speech and for monitoring our own speech. Because this theory states that monitoring is localized in the perception apparatus, it would predict perception-specific effects on monitoring of internal speech. Evidence for such perception-specific effects comes from studies using the phoneme monitoring task. In this task, participants monitor their own internal speech for a pre-specified target phoneme. For instance, Özdemir, Roelofs, and Levelt (2007) presented participants with pictured objects and the task of the participant was to indicate whether the picture name contained a particular phoneme. In these monitoring tasks effects of serial position in the monitoring of internal speech have been demonstrated. The monitoring latencies increased from the beginning to the end of the word; i.e., phonemes at the beginning of a word were detected faster than word-medial and word-final phonemes (Wheeldon & Levelt, 1995; Özdemir et al., 2007). Apparently, a word's phonemes are made available from left to right which results in activation of the Speech-Comprehension System in a serial manner, as is the case with the processing of overt speech. So, the left-to-right time course can be taken to indicate that monitoring of internal speech is indeed localized in the perception apparatus, as proposed by Levelt's perceptual loop theory.

Monitoring in language perception

In addition to errors of language production, we also make errors in perception, for example when we misread a word or misunderstand a speaker. Many of the studies of spontaneous misperceptions or 'slips of the ear' include a great number of

examples of such errors (Cutler & Butterfield, 1992). For instance, the sentence *'She's a must to avoid.'* which was heard as *'She's a muscular boy'.* In an amusing historical anecdote, a British officer is reputed to have passed along the message, *'Send reinforcements, we're going to advance',* which was famously misinterpreted as, *'Send three and four pence, we're going to a dance'.* Such unexpected interpretations can of course represent true message elements the other speaker wants to convey (e.g., they might actually have been going to a dance). On the other hand, they could also result from perceptual processing errors. In the present thesis, we propose that we also monitor for perceptual errors, to distinguish between these two possibilities. Such a distinction is important because it determines, for example, whether military leaders will send pocket money, or additional troops!

One proposal put forward in this thesis is that monitoring in language perception serves to distinguish between true message elements and processing errors. But, in the case of perceived speech, errors cannot be observed directly in the way that they can be detected in spoken language (as in the case of an incorrectly uttered word). It is also impossible to have access to the internal process responsible for the error. So the question becomes: how can language-users know that a perceptual processing error has occurred? Or in other words; how can perceptual errors be detected by the listener? The only cue that could tell the listener that something went wrong consists of a conflict between what one expects to hear and what one actually hears. If you think you hear something about a muscular boy or a dance but that's not relevant to the conversation then the unexpectedness of that tells you that something went wrong. Accordingly, in this thesis we propose that it is the violation of the expectation or the conflict between what you perceive and what you expect which signals a possible processing error in language perception. This conflict would trigger a monitoring response to check for the possibility of such an error.

This dissertation contains three event-related potential (ERP) studies that support the notion that a very strong conflict between what an average language user expects and what is actually presented elicits a late positivity in the electroencephalogram (EEG). This late positivity is proposed to reflect reprocessing meant to repair the possible perceptual processing error. In the next part, a brief introduction to the technique of event-related brain potentials is given, accompanied by an overview of the main findings in the ERP literature that served as the starting point for the ERP work reported in this thesis.

Event Related Brain Potentials (ERPs)

Variations in the brain's electrical activity over time can be measured by means of electrodes attached on the scalp. The record of this brain activity is called the electroencephalogram (EEG). The electrical activity depicted by the EEG is the result of simultaneous post-synaptic activity within neocortical pyramidal neurons. The brain's spontaneous electrical activity is referred to as background EEG. Changes in background EEG activity reflect large changes in the general state of the subject, for example as a function of being awake or asleep. In contrast, event-related brain potentials (ERPs) represent small changes in the electrical activity of the brain that are elicited by some sensory, cognitive or motor event (see e.g., Coles & Rugg, 1995). An ERP is much smaller in amplitude (5-10 μV) in comparison to the spontaneous background EEG (50-100 μV). Therefore, ERPs to a particular event are usually not visible in the raw EEG. In order to extract the ERP from the background EEG we have to average the EEG over repeated presentations of 'similar' stimuli. These stimulus events are not precisely the same, but are variations within a specific stimulus event-class (e.g. a set of nouns matched for frequency and length). Fluctuations in electrical activity generated by neurons which are not involved in processing the stimuli of interest will be random with respect to the time of stimulus onset and thus average each other out. This leaves a record of the event-related activity (ERP) time-locked to the presented stimuli. The number of trials needed for a reliable ERP average is a function of the amplitude of the ERP component of interest. The smaller the component, the more trials are needed to extract it from the spontaneous EEG, but 25 observations per condition should be considered as a minimum (e.g., Kutas & Van Petten, 1994; Kutas, Van Petten, & Kluender, 2006).

The ERP measurement is differential, which means that we place an active electrode over the brain area of interest and subtract from it the activity of a passive, i.e., electrically silent area, the so-called reference electrode. Most often the mastoids or the earlobes are chosen as reference, because these are considered to be little

affected by the electrical activity of the brain. The amplitude of the peaks in an ERP is thus the difference in electrical charge between the active electrode and the reference electrode. Since eye movements seriously distort the EEG recording, it is necessary to record the vertical as well as the horizontal eye movements during the experiments. Before analysing the ERP data, trials with ocular artefacts above a critical value (usually about 100 μV) as well as trials contaminated by other biological artefacts (e.g., muscular activity, large electro-cardiographic potentials, and changes in skin conductance) have to be rejected for further analysis.

An ERP waveform includes a series of positive and negative voltage peaks. One should be aware of the fact that the term ERP component is generally not the same thing as a peak or trough in the ERP waveform, but that it refers to a more theoretical concept. Donchin, Ritter, and McCallum (1978) give an operational definition of an ERP component. According to this view, a component is a part of the waveform with a circumscribed scalp distribution (alluding to the underlying neural configuration) and a circumscribed relationship to an experimental variable or to a combination of experimental variables (alluding to the cognitive function served by the activity of this configuration). A general distinction is made between exogenous and endogenous ERP components. The exogenous or stimulus-bound components occur within less than 100 ms after the presentation of a stimulus and represent the obligatory brain response to the stimulus. The amplitude and form of these early components is largely determined by physical parameters of the stimulus, such as intensity, frequency, and rate of presentation. Exogenous components are generally impervious to a subject's state of alertness or attentiveness. Relevant for language research are the endogenous components with latencies usually beyond 100 ms. The endogenous components are relatively insensitive to variations in physical stimulus characteristics, but are mainly affected by the cognitive aspects of stimulus processing like the level of attention for a stimulus or instructions. Endogenous ERP components are not evoked by a stimulus per se but are elicited by the cognitive processing of that type of stimuli. The same stimuli may or may not evoke a particular endogenous component depending upon how the subject 'chooses' to process them.

The ERP component nomenclature is usually based on the polarity and peak latency measured from stimulus onset. For instance, the P300 stands for a positive peak with a latency of about 300 ms after onset of the critical stimulus. On occasion, the peaks are labelled by their polarity and ordinal position in the waveform (eg., N1, P1, N2). And sometimes, the component is named after its assumed cognitive function; an example is the 'Syntactic Positive Shift (SPS)'.

ERP correlate of semantics

In 1980 Kutas and Hillyard published a landmark study in which they reported that semantically inappropriate words (e.g., 'He spread the warm bread with *socks.*') elicited a large negative ERP component with a peak latency of about 400 ms (hence the N400 component), relative to semantically appropriate words (e.g., 'He spread the warm bread with *butter.'*). In subsequent studies Kutas and colleagues established that all open class words elicit an N400. In addition, the amplitude of the negative wave to appropriate sentence-final words was smaller compared to the anomalous sentence endings. This modulation of N400 amplitude by semantic context is referred to as the *N400 effect*. Important for the present purposes is that several studies have shown that an N400 effect can also be elicited by semantically correct sentences; that is, N400 effects have also been reported to correct but less expected words, as in 'He mailed the letter without a *thought*.' compared to 'He mailed the letter without a *stamp*.' (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). Specifically it has been shown that N400 amplitude is inversely related to cloze probability [1], i.e., the higher the cloze probability the smaller the N400. More recently, Hagoort, Hald, Bastiaansen, and Petersson (2004) have shown that words that violate real world-expectations such as 'Dutch trains are *white* …' elicit a larger N400 amplitude compared to words that are consistent with such expectations, as in 'Dutch trains are *yellow* …'. A well established view on the N400 is that it reflects the ease with which a word is integrated into the current context, be this a single word (e.g., Holcomb, 1993; Chwilla, Brown, & Hagoort, 1995; Chwilla, Hagoort, & Brown, 1998; Chwilla, Kolk, & Mulder, 2000), a sentence

[1] The amount of contextual constraint imposed by a sentence fragment can be assessed by means of a cloze probability test. The cloze probability refers to the proportion of subjects that fill in a particular word as the best completion of a sentence fragment (cf. Taylor, 1953).

context (e.g., Friederici, 1995; Van Petten & Kutas, 1990) or a discourse context (St. George, Mannes, & Hofman, 1994; Nieuwland & Van Berkum, 2005; but see for example for a different view of the N400 in terms of lexical access Deacon, Hewitt, Yang, and Nagata, 2000).

ERP correlate of syntax

One of the most robust findings is that manipulations of syntactic structure elicit a late positive shift starting at about 500 ms and typically extending up to at least 800 ms after critical word onset. This positivity is commonly referred to as P600. A P600 was firstly reported by Osterhout and Holcomb (1992) after so-called garden path sentences. In these garden path sentences there is a strong preference for a particular structural analysis, which turns out to be the wrong one later on in the sentence. For example, readers of the sentence 'The broker persuaded *to* sell the stock.' initially assume that the sentence is about the broker persuading someone, but when reading the word 'to' they realize that the sentence is about the broker being persuaded. A similar P600 effect has also been observed in response to several sorts of grammaticality violations such as syntactic agreement violations, e.g., 'The spoilt child *throw* the toys on the floor.' and word order violations, e.g., 'the expensive very *tulip*.' (e.g., Hagoort, Brown, & Groothusen, 1993). In addition, P600 effects are reported in some grammatical but very complex sentences, as compared to less complex controls (Kaan, Harris, Gibson, & Holcomb, 2000). On the basis of these results, the P600 effect was assumed to reflect syntactic reanalysis (Friederici, 1995; Osterhout, Holcomb, & Swinney, 1994), syntactic processing as such (Hagoort et al., 1993) or syntactic integration difficulty (Kaan et al., 2000)

However, more recently, P600 effects have been observed after different types of semantic anomalies in syntactically unambiguous sentences; this has been shown in different languages (Kolk, Chwilla, van Herten, & Oor, 2003; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003, Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Van Herten, Kolk, & Chwilla, 2005; Van Herten, Chwilla, & Kolk, 2006). For instance, Kuperberg et al. (2003) observed a P600 effect after anomalous verbs in simple active sentences, for example after 'eat' in the sentence 'Every morning at breakfast the eggs would *eat* …'. The occurrence of these P600 effects to semantic anomalies in syntactically unambiguous sentences challenges the view that semantic versus syntactic anomalies indeed elicit qualitatively different ERP patterns.

Still, to date, the dominant view is that the P600 effect reflects a process of structural reanalysis in language perception. Researchers including Friederici (1995), Kuperberg et al. (2003), and Kim and Osterhout (2005) have proposed a process of syntactic reanalysis in such cases as the semantically anomalous sentences described above. In many types of semantic anomalies that evoke a P600 rather than an N400, there was a strong semantic-thematic 'attraction' (Kim & Osterhout, 2005) or a potential semantic-thematic 'fit' (Kuperberg et al., 2003) between the critical verb and its preceding argument. Specifically, the semantic anomalies were verb-argument semantic violations in which the arguments could have occupied alternative thematic roles which are more plausible. Many of these anomalous sentences were, therefore, repairable by reassigning the thematic roles of the critical verb's arguments to the more plausible thematic roles. For example, sentences like 'Every morning at breakfast the eggs would *eat* …', would make sense if the reader would reassign the agent role of 'eggs' to a theme role, such as in 'Every morning at breakfast the eggs would *be eaten* …'. So, the P600 effects observed after semantic anomalies with a strong semantic fit or attraction is proposed to represent syntactic reanalysis; involving an online attempt to structurally repair a sentence by reassigning thematic roles. Accordingly, these researchers put forward that the function of the P600 effect after semantic anomalies is to structurally repair a sentence by reassigning thematic roles.

In this dissertation, we take the same view that the P600 reflects sentence reanalysis. However, we claim that this sentence reanalysis has a more general function than syntactic restructuring. That is, we put forward that the function of the sentence reanalysis is to check whether the inconsistency arises from a processing error due to misreading. It is like asking yourself: Did I read that correctly? As described above, in the case of language perception, an inconsistency can have two sources. It can be real, in the sense that an unexpected event has indeed occurred (e.g., that the girl *is* a muscular boy or that the eggs are eating or that the man bites the dog). On the other hand, it can also stem from a processing error. To prevent integration of erroneous information into the discourse, the reader will

generally monitor the correctness of his or her analysis in case of a conflict between a highly expected representation and an unexpected representation.

Structure of the thesis

In this thesis, we propose that the P600 reflects reprocessing with the more *general* function of checking whether the initial sentence processing was correct. Because different linguistic elements can be misperceived, we predict that the P600 effect could reflect reprocessing at a number of different linguistic levels. To test this hypothesis in the three experimental chapters, we induced conflicts at different levels of the linguistic system. In Chapter 2 a conflict was produced at the *sentence* level, in Chapter 3 we induced a conflict at the *word* level and in Chapter 4 a conflict was elicited at the *conceptual* level. The underlying rationale of this experimental approach was to investigate monitoring at three different linguistic levels. The chapters are written in such a way that they can be read independently. An unavoidable consequence is that some overlap exists between the introduction and the method sections. In the following paragraphs, possible conflicts at the three different linguistic levels and the resulting experiments are introduced.

Chapter 2: Conflict at the sentence level

A monitoring process at the sentence perception level was firstly described by Kolk and colleagues in 2003. Because this paper is of direct relevance for the research described in **Chapter 2**, we briefly sketch the rationale and main results of this paper. Kolk and colleagues assume that, in sentence perception, simple processing heuristics are used in addition to syntactic algorithms, which together determine the final interpretation of the sentence (see also, Bever, 1970; Ferreira, Ferraro, & Bailey, 2002; Ferreira, 2003). Heuristics can be regarded as 'rules of thumb': highly economical strategies that are generally but not always effective in extracting meaning. One proposed heuristic is the plausibility strategy. According to this semantic heuristic, readers depend heavily on their knowledge of the meaning of individual content words, which provide a strong basis for the most plausible interpretation. The syntactic algorithm on the other hand involves an algorithmic analysis of the syntactic structure of the sentence; this analysis is time-consuming but always comes up with the correct interpretation. Now the question arises of how these heuristics and parsing systems are coordinated. One possibility is a cascade-like model in which heuristics constrain the initial hypothesized search space of the subsequent algorithmic parser (Townsend & Bever, 2001; Bever, Sanz, & Townsend, 1998). If this parser has time to finish all its computations, it will output the complete and correct interpretation of the sentence. As Townsend and Bever (2001) state 'semantics propose, syntax disposes.' (p.271). Another possibility is that the two routes operate in parallel and are largely independent from one another. Kolk and colleagues (2003) and Van Herten, Chwilla, and Kolk (2006) argue in favour of this view and point to the dual route model of reading aloud, in which there is similar parallel processing along two independent routes (e.g., Coltheart, Curtis, Atkins, & Haller, 1993). If there is such parallel processing in the case of sentence interpretation, it would be possible for the two routes to lead to conflicting outcomes. Kolk et al. (2003) propose that it is just such a conflict that triggers a monitoring response after semantic reversal anomalies. These semantic anomalies were formed by exchanging the subject and object of semantically acceptable sentences, as for example in 'De vos die op de stropers *joeg* sloop door het bos.' (Paraphrase: 'The fox that *hunted*[singular] the poachers stalked through the woods.'). It is clear that in this sentence, plausibility heuristics and syntactic algorithms produce different thematic interpretations. Whereas the plausibility heuristic leads to the interpretation that *the poachers hunted the foxes*, the parsing routines lead to the interpretation that *the foxes hunted the poachers*. This conflict between the semantically plausible, highly expected thematic interpretation and the implausible thematic interpretation makes it necessary for the brain to re-attend the unexpected linguistic unit to verify its veridicality. Kolk et al. (2003) observed a P600 effect in response to these semantic reversal anomalies which is assumed to reflect this general check for processing errors.

The goal of the ERP experiment described in Chapter 2 was to test whether the P600 effect after semantic reversal anomalies reported by Kolk et al. (2003) does indeed reflect a control process triggered by a conflict at the sentence level. As described above, the P600 effect reported by Kolk et al. (2003) is assumed to be triggered by a conflict between the outcome of a plausibility heuristic with that of a parsing routine. From this, we predicted that if this P600 effect reflects a control

process to check for possible processing errors, it should be reduced when the discrepancy between the thematic interpretations proposed by the semantic heuristic and the syntactic parse is diminished. To test this hypothesis, we used the same stimulus materials that were used in the study of Kolk and colleagues (2003) but we directed the participants' attention to the syntactic level. We explicitly informed the participants that semantic reversals had been intentionally constructed and that they should not be misled by word meanings, but attend to the structure of the sentence. This focus-on-syntax instruction created a context for participants wherein anomalous sentences were expected, the anomalies were therefore less likely to be interpreted as possible processing errors. Consequently, this instruction should reduce the discrepancy between the thematic interpretations proposed by the semantic heuristic and the syntactic parse. If Kolk and colleagues' proposal (2003) that the P600 effect after semantic reversal anomalies is based on a control operation triggered by a conflict at the sentence level is correct, then this focus-on-syntax instruction should lead to a decrease in error rates and a reduction or even disappearance of the P600 effect.

The main conclusion of this chapter is that a substantial part of the P600 effect after semantic reversals can be accounted for by the control operation that is triggered by a conflict between two incompatible interpretations at the sentence level.

Chapter 2 has been published as: Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. Biological Psychology, 75, 8-18.

Chapter 3: Conflict at the word level

As described above, the P600 effect is proposed to reflect reprocessing with the more *general* function of checking for processing errors. As a consequence, it could involve reprocessing at a number of linguistic levels. The aim of the ERP experiment reported in **Chapter 3** was to test whether a conflict between two representations can also trigger a monitoring response at the word level.

It is relevant to point out in this context that Münte, Heinze, Matzke, Wieringa, and Johannes (1998) observed a P600 effect after orthographic anomalies in a study with German-speaking participants, as for example in response to 'Die Hexe benutzte ihren *Behsen*, um zum Wald zu fliegen.' (Literal translation: 'The witch used her *broome* to fly to the forest.'). This P600 effect could stem from a conflict, similar to what we proposed for the semantic anomalies. There is a strong tendency to accept the word '*Behsen*'. First, because it is semantically highly expected and second, because the phonological form of the word confirms this expectation and makes it maximally strong. On the other hand, there will be a strong tendency to reject the word since the orthographic form does not fit the phonological form: the word is misspelled. The source of the conflict is very different from what we saw in the case of the semantic anomalies. Here, the tendency to accept stemmed from the fact that a plausibility heuristic indicated a highly plausible interpretation based on world knowledge and the tendency to reject stemmed from the fact that the regular parse indicated a highly implausible interpretation.

In our test of the monitoring hypothesis at the word level, we followed Münte et al. (1998) by presenting pseudohomophones in high-cloze contexts. The critical lexical item was either spelled correctly or was a pseudohomophone derived from the expected word and phonetically similar to the expected word, for example 'In die bibliotheek lenen scholieren *boekun* om mee naar huis te nemen.' (Paraphrase: 'In that library the pupils borrow *bouks* to take home.'). What we added in this study was a low-cloze condition, for example 'De kussens zijn opgevuld met *boekun* waardoor ze hard aanvoelen.' (Paraphrase: 'The pillows are stuffed with *bouks* which make them feel hard.')

For the high-cloze context, we predicted the system to anticipate that a particular word will occur, and then start up a monitoring process when a misspelled word that is phonologically identical and orthographically similar to the highly expected word is actually presented. This monitoring process at the word level is expected to elicit a P600. For the low-cloze items, the lexical items from which the pseudohomophones are derived are not highly expected and thus should not elicit a conflict between the expected and actually presented lexical item. Consequently, no monitoring process and hence no P600 effect were expected to occur.

The main conclusion of this chapter is that a conflict between a strong tendency to accept and a tendency to reject a representation can trigger a monitoring response at the word level which is reflected by a P600 effect.

Chapter 3 has been published as: Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Chapter 4: Conflict at the conceptual level

The purpose of the experiment described in **Chapter 4** was to test whether in addition to a process of monitoring for errors at the sentence level (Chapter 2) and at the word level (Chapter 3), there is also a monitoring process after sentence-picture mismatches at the conceptual level.

A sentence-picture matching task was also used in investigating on-line thematic role assignment by Wassenaar and Hagoort (2007). They presented subjects with a picture that was followed by a syntactically correct sentence. The thematic roles of the sentence either matched or mismatched the thematic roles displayed in the picture. For example, in the mismatch condition, after presentation of a picture in which a woman pushes a man in a wheelchair, the following sentence was presented 'De lange man op dit plaatje *duwt* de jonge vrouw' (Literal translation: 'The tall man on this picture *pushes* the young woman.'). A P600 effect was observed after the sentences that did not fit the depicted thematic roles. Wassenaar and Hagoort (2007) propose that in these sentences the role assignment based on the picture interferes with the role assignment based on the sentence; since the picture indicates one role assignment, it is difficult to assign the reverse thematic roles in the sentence. This interference hypothesis proposes the size of the P600 effect to vary as a function of how effortful the assignment process is. The Monitoring Theory accounts for the P600 effect in a different way, by proposing that it reflects a monitoring response triggered by the conflict between the predicted thematic roles on the basis of the picture representation and the thematic roles of the sentence.

In the experiment described in Chapter 4, we presented syntactically correct and unambiguous sentences which correctly or incorrectly describe a preceding picture with geometrical figures. In particular, a picture of a spatial array was shown, for instance of a square in front of a triangle. Then a sentence was presented which could either match or mismatch with the picture, depending on the preposition being used. As for example in the following picture-sentence mismatch: □ △
'De driehoek staat *boven* het vierkant.' (Paraphrase: 'The triangle stands *above* the square.')

Our prediction was that the picture-sentence mismatches would create a conflict between the conceptual representation emanating from the picture and the conceptual representation derived from the sentence and would therefore elicit a P600. By employing locative relationships, we avoided the involvement of thematic role assignment and thereby a possible effect of role assignment interference, as proposed by Wassenaar and Hagoort (2007). This is because thematic roles refer to actions, in particular to the way different entities are involved in such actions. So, the predicted P600 effect for sentence-picture mismatches could not be related to thematic role interference.

The main conclusion of this chapter is that a conflict between a highly expected representation and an unexpected representation can also trigger monitoring at the conceptual level reflected by a P600 effect.

Chapter 4 has been published as: Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk, H.H.J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. Neuropsychologia, 46, 967-982.

Chapter 5 provides a summary of the main results of the three experimental chapters. We conclude with a number of frequently asked questions concerning the main findings of the present thesis and the Monitoring Theory.

References

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362). New York: Wiley.

Bever, T.G., Sanz, M., & Townsend, D. J. (1998). The emperor's psycholinguistics. *Journal of Psycholinguistic Research,* 27, 261-284.

Chwilla, D.J., Brown, C.M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology,* 32, 274-285.

Chwilla, D.J., Hagoort, P., & Brown, C.M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology,* 51A, 531-560.

Chwilla, D.J., Kolk, H.H.J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language,* 42, 314-341.

Coltheart, M., Curtis, B., Atkins. P., & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed processing approaches. *Psychological Review,* 100, 589-608.

Coles, M.G.H., & Rugg, M.D. (1995). Event-related brain potentials : An introduction. In M.D. Rugg & M.G.H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1-26). New York: Oxford University Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language,* 31, 218-236.

Deacon, D., Hewitt, S., Yang, C.M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research,* 9, 137-146.

Donchin, E., Ritter, W., & McCallum, W.C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds), *Event-Related Brain Potentials in Man*, (pp 349-441). New York: Academic.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology,* 47, 164-203.

Ferreira, F., Ferraro, V., & Bailey, K.G.D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science,* 11, 11-15.

Friederici, A.D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language,* 50, 259-281.

Garnham, A., Shillcock, R., Brown, G., Mill, A., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics,* 19, 805-817.

Holcomb, P.J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology,* 30, 47-61.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes,* 8, 439-483.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science,* 304 (5669), 438-441.

Hoeks, J.C.J., Stowe, L.A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research,* 19, 59-73.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes,* 15, 159-201.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language,* 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language,* 85,1-36.

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research,* 17, 117-129.

Kuperberg, G.R., Kreher, D.A., Sitnikova, T., Caplan, D.N., & Holcomb, P.J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language,* 100, 223-237.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science,* 207, 203–205.

Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature,* 307, 161–163.

Kutas, M., Lindamood, T., & Hillyard, S.A. (1984). Word expectancy and event-related potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–238). Hillsdale, NJ: Erlbaum Press.

Kutas, M., & Van Petten, C.K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Kutas, M., Van Petten, C.K.F., & Kluender, R. (2006). Psycholinguistics electrified II. Event-related brain potential investigations. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics, 2nd edition.* New York, NY: Elsevier Press.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levelt, W.J.M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics,* 205-218.

Münte, T.F., Heinze H.-J., Matzke M., Wieringa, B.M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia,* 217-226.

Nieuwland, M.S., & Van Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research,* 24, 691-701.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language,* 31, 785-806.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 20(4), 786-803.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition,* 105, 457-465.

Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. *Syntax and Semantics,* 12, 261-286.

Stuss, D.T., & Benson, D.F. (1986). The frontal lobes. New York: Raven Press.

St. George, M., Mannes, S., & Hoffman, J.E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6, 70-83.

Taylor, W.L. (1953). “Cloze” procedure: A new tool for measuring readability. *Journalism Quarterly,* 30, 415-417.

Townsend, D.J., & Bever, T. G. (2001). Sentence comprehension: *The integration of habits and rules.* Cambridge, MA: The MIT Press.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word-frequency in event-related brain potentials. *Memory and Cognition,* 18, 380-393.

Van Herten, M., Chwilla, D.J., & Kolk, H.H.J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitve Neuroscience,* 18, 1181-1197.

Van Herten, M., Kolk, H.H.J., & Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18.

Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk H.J. (2008) Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967-982

Wassenaar, M., & Hagoort, P. (2007) Thematic role assignment in patients with Broca's aphasia: Sentence-picture matching electrified. *Neuropsyhcologia*, 45, 716-740.

Wheeldon, L.R., & Levelt, W.J.M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 226-244.

Chapter 2

The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times[1]

[1] This chapter has been published as Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18.

Abstract

Semantic anomalies like '*the fox that hunted the poacher*' elicit P600 effects. Kolk, Chwilla, Van Herten and Oor (2003) proposed that this P600 effect is triggered by a conflict between the outcome of a lexical strategy with that of the parsing routine. Specifically, when the lexical strategy indicates that the poacher hunted the fox, the full parse leads to the conclusion that the fox was the one who did the hunting. We tested this hypothesis by replicating the study cited above but manipulating the context by means of instruction. Participants were informed that semantic anomalies were created on purpose and that they should not be misled by these anomalies but instead focus on syntax or sentence structure. This instruction led to a strong reduction in P600 effect. This result supports the view that expectations play an important role in the generation of P600 effects to semantic anomalies, as proposed by Kolk and colleagues (2003).

Introduction

Under normal circumstances we rarely experience difficulties in understanding sentences in our mother tongue. People comprehend sentences rapidly and without conscious effort. However, numerous problems have to be resolved during the process of understanding a sentence. Sentences are not just strings of words, linked together in a random way. Word by word, information from several sources becomes available that readers and listeners must use, specifically: pragmatic, semantic and syntactic information. Theories about the way in which these sorts of information are represented, retrieved and combined during language processing have urged different lines of psycholinguistic research during the last decade. In the 1960s, most psycholinguistics shared Chomsky's view that syntactic processing proceeds independently of semantic information. This view is exemplified by Chomsky's example sentences (Chomsky, 1957, p.15): 'Colorless green ideas sleep furiously.' and 'Furiously sleep ideas green colorless.' These two sentences are equally absurd, but one will recognize that only the former is syntactically correct. Chomsky (1957, p.15) says: 'The notion "grammatical" cannot be identified with "meaningful" or "significant" in any semantic sense.' Fodor (1983) assimilated the concept of independent syntactic processing into the general idea that syntactic processing is modular. Modular processes are algorithmic, autonomous and operate bottom-up. These processes are not guided by world knowledge, beliefs or expectations. So, according to Fodor, syntactic processes are not guided by the meaning or plausibility of the sentence.

The so called syntax-first theories embody the above-mentioned claim that the syntactic module is restricted to the domain of syntactic information and is immune to non-syntactic information. In this module, an initial commitment is made to a single syntactic structure of the sentence as it has developed so far, on the basis of syntactic information alone. Semantic and pragmatic information is represented in another module and is only activated at a later processing stage (e.g., Frazier & Fodor, 1978; Ferreira & Clifton, 1986).

Constraint satisfaction models have been proposed as a competitor of syntax-first models. According to these models, information from all sources interacts continuously during the formation of an internal interpretation. That is, all kinds of information provided by the incoming words (e.g., context, discourse and semantic information) can jointly affect the activation of different syntactic alternatives (e.g., McDonald, Maryellen, Pearlmutter, & Seidenberg, 1994; St John & McClelland, 1990; Trueswell & Tanenhaus, 1994).

Ever since Fodor (1983) argued that parsing is basically a reflex, most models of sentence comprehension have assumed that an interpretation must be based on an initial syntactic structure, even if the activation level of that structure can be influenced by nonsyntactic sources of information. This implies that, according to both syntax-first models and constraint satisfaction models, every interpretation is complete, detailed and accurate and can not be based on shallow, inaccurate or incomplete processing. However, the meaning people derive from a sentence is often not a reflection of its true content. Hence, what current models are missing is an architectural component that can explain cases in which people use strategies or engage in heuristic processing of sentences that, may then result in an inaccurate interpretation (Ferreira, Bailey, & Ferraro, 2002; Ferreira, 2003).[2]

Researchers in other domains of cognition, such as decision making and reasoning, have argued that human behavior is at least partially driven by heuristic processing (Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & ABC Research Group, 1999). Gigerenzer and his colleagues have proposed the idea of bounded rationality. They argue that humans often find themselves in situations which force them to make inferences about the world under limited time, knowledge, and computational power. Models of rational inference do not take these limitations into account and are therefore unrealistic. They show that simple heuristics can match or even outperform classical models of rational inference. These simple heuristics are frugal because they exploit the structure of environments and only require the use of a small proportion of the available information. As a consequence, they are thought to be fast because information search is less computationally demanding.

[2] With heuristics we mean that in some situations language users do not take all relevant information into account, syntactic as well as semantic information, but only a specific part. As we will see below, this part may either be word order or the meaning of the set of lexical items.

How might heuristics be used during language processing? One proposed heuristic is referred to by Bever (1970) as ‘strategy C’, and by Ferreira and colleagues (2002) as ‘the plausibility strategy’ and will be referred to below as ‘the lexical strategy’. This strategy is a semantic heuristic which states that readers depend heavily on their knowledge of the meaning of individual content words, which provide a strong basis for the most plausible interpretation. Hence, according to this strategy, readers depend on schemas in long-term memory or world knowledge. Evidence for the existence of such a bias comes from work with aphasic patients (Saffran, Schwartz, & Linebarger, 1989). In particular, these patients have great difficulty rejecting sentences like ‘*the painting disliked the artist*’, even though this is a simple active construction, a sentence type they have little difficulty with in other tests. Another heuristic is the canonical ‘word order strategy’ (Townsend & Bever, 2001; Bever, Sanz, & Townsend, 1998). This heuristic says that a noun phrase preceding a verb is taken as the subject of that verb. Similarly, a noun phrase following a verb is taken to be the object of the verb. Since this order is present in the majority of sentences in the English language, this is a very advantageous strategy.

Using a plausibility strategy implies a process of integration: one takes a set of content words and attempts to fit these words into a meaningful whole. Now, precisely such a process of meaning integration is what appears to set off a well-known language-relevant ERP component: the N400. The N400 is a negative voltage peak that reaches its maximum amplitude around 400 ms after the onset of open class words (for reviews: Kutas & Van Petten, 1994; Kutas & Federmeier, 2000; Kutas & Schmitt, 2003). Pairs of words that do not fit well together semantically - e.g. 'cat' and 'rose' - elicit larger N400 amplitudes than pairs of words that do -e.g., 'cat' and 'dog' - (e.g., Chwilla, Brown, & Hagoort, 1995; Chwilla, Hagoort, & Brown, 1998). That this effect truly reflects meaning integration was further confirmed by the results of a study of Chwilla, Kolk, and Mulder (2000). These authors found that indirect semantic relationships between words (e.g., Prime: ‘lion’, Target: ‘stripes’, Mediator [not presented]: ‘tiger’) elicited an N400 effect only if these indirect relationships were the strongest relationships in a list. In contrast, if the list also contained directly related pairs (e.g., ‘girl’ and ‘boy’), no effect of indirect relationships was found. This led the authors to conclude that the participants attempted to find semantic coherence in a pair of words at the highest level of coherence that the material permits.

The plausibility strategy described above refers to a process of meaning integration between a set of content words that occurs in a sentence. Therefore, the question arises whether the process of meaning integration for pairs of isolated words that is reflected by the N400 is similar to the integration process also reflected by the N400 within sentences. This question has been investigated by Kutas (1993) by comparing the N400 effects to word pairs and sentences in the same group of subjects. The outcome was that the N400 relatedness effect does not depend upon whether the words occur in a list or in a sentence: as the latency, amplitude and overall shape of the N400 effects in the two conditions were very similar. It therefore seems likely that a process of meaning integration as we described above, in which participants attempt to find maximal coherence between content words, also takes place within sentences. This is exactly what the plausibility strategy embodies.

The evidence reviewed above suggests that a comprehensive theory of language comprehension should assume that simple processing heuristics are used in addition to syntactic algorithms. The question remains as to how these heuristics and parsing systems are coordinated. According to the Late Assignment of Syntax Theory (LAST) (Townsend & Bever, 2001; Bever et al. 1998), the initial semantic analysis of sentences proceeds on the basis of statistically sensitive perceptual strategies or heuristics (the authors only mention the word-order strategy here, but the same reasoning could be applied to the plausibility strategy). Townsend and Bever (2001) refer to this analysis as a "pseudo-parse", because the comprehension mechanism does not operate as the mere mechanical application of syntactic categories and frames from left to right. Instead, the pseudo-parse is a probabilistic analysis of meaning and form that proposes a likely candidate meaning or conceptual structure. So, the pseudo-parser uses heuristics to create a preliminary hypothesis or 'best-guess' about the input. In addition, a more time-consuming algorithmic analysis proposes a candidate real syntax. This true parser uses the preliminary hypothesis to constrain its initial hypothesized search space. If this algorithmic parser has time to finish all its computations, it will output the complete and correct interpretation of the sentence. Townsend and Bever (2001) state: 'semantics propose, syntax disposes', (pp 271). Ferreira (2003) also proposes a dual route model and argues that the language comprehension system uses a combination of heuristics and syntactic algorithms.

In most sentences, heuristics and syntactic algorithms will produce the same thematic interpretation. However, in particular sentences, the two routes may produce conflicting results. Such a conflict may underlie the ERP findings, obtained by Kolk, Chwilla, Van Herten and Oor (2003). They used semantic reversal anomalies which were formed by exchanging the subject and object of semantically acceptable sentences such as (1).

(1) De vos die op de stropers joeg sloop door het bos (original).
The fox that at the poachers hunted [singular] stalked through the woods (literal translation).
The fox that hunted [singular] the poachers stalked through the woods (paraphrase).

It is clear that in these sentences lexical strategies and parsing routines produce different thematic interpretations. Whereas the lexical strategy leads to the interpretation that poachers are hunting foxes, the parsing routines lead to the interpretation that the foxes are hunting the poachers. Although the latter interpretation is not entirely impossible it represents a highly unlikely event based on world knowledge.

Kolk and colleagues (2003) observed a P600 effect- a late positive potential starting at about 600 ms after the onset of a target word- and not an N400 effect to these semantic reversal anomalies. This is consistent with recent findings from other researchers who, despite differences in sentence material and language (English and Dutch), observed a P600 effect in the absence of an N400 effect to semantically implausible sentences relative to their plausible counterparts (Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Van Herten, Kolk, & Chwilla, 2005, and see also: Vissers, Chwilla, & Kolk, 2006). These results seem to challenge the view that P600 effects are reliably elicited by syntactic anomalies whereas N400 effects are elicited by semantic anomalies. [3] How can one account

[3] For a discussion of different views on the occurrence of a P600 effect after semantic anomalies, see Van Herten et al. 2005.

for these unexpected results and how does it relate to the presence of a conflict between algorithmic and heuristic processing routes?

Kolk and colleagues (2003) proposed that the P600 effect reflected the conflict between the thematic interpretations whereas the absence of the N400 effect was due to the fact that the 'lexical' interpretation was the same for plausible and implausible sentences. According to the above described semantic integration view of the N400, the attempt to integrate sets of individual content words into one coherent meaning, is reflected in the N400. Since the word level integration process does not have difficulty integrating the words of the semantic reversals in the study of Kolk et al. (2003) into a coherent meaning (the lexical items in both conditions are the same; the fox and the hunters) readers initially do not notice the anomaly. Hence, no N400 effect was elicited.

Why would a P600 effect follow a conflict between thematic interpretations? Kolk and colleagues (2003) argued that the language comprehension system attempts to resolve the conflict by reprocessing the sentence to check the memory trace of the input sentence for possible processing errors. In particular, the mismatch between the semantically plausible, highly expected (based on world knowledge) thematic interpretation and the implausible thematic interpretation makes it necessary for the brain to re-attend the unexpected linguistic unit to check upon its veridicality. After all, an inconsistency can have two sources. It can be real, in the sense that an unexpected event has indeed occurred (e.g. man bites dog). On the other hand, it can also stem from a processing error. To prevent integration of erroneous information into the current discourse, the reader will generally check upon the correctness of his or her analysis in case of a conflict. This explains the occurrence of P600 effects to semantic reversal anomalies.

The levels of processing framework has been used to further determine the processing nature of the N400 component (e.g., Chwilla, Brown, & Hagoort 1995; Besson, Fischler, Boaz, & Raney, 1992; Kutas, & Hillyard, 1989) and the P600 component (e.g., Gunter and Friederici, 1999; Gunter, Stowe, & Mulder, 1997; Hahne, & Friederici, 1997). According to this framework, different task demands are assumed to result in different levels of processing (Craik & Lockhart, 1972) during word processing. For instance, Gunter and Friederici (1999) used a physical task in

the shallow processing condition, in which participants had to judge whether a word in a sentence was printed in upper case (a shallow level of processing). The deep processing condition was a grammatical task in which subjects had to judge sentence-final syntactic errors. This task manipulation was effective in modulating the amplitude of P600; the physical judgment task greatly attenuated or eliminated the N400 and P600 following incorrect verb inflections compared to the grammatical judgment task. This was taken to indicate that the P600 mainly reflects controlled syntactic processing.

In the present paper we used a similar approach to investigate whether the P600 to semantic reversal anomalies is modulated by instruction. Rather than drawing the participants' attention to the physical level as Gunter and Friederici (1999) did, we drew the participants' attention to the syntactic level. More specifically, the aim of the present experiment was to test the hypothesis by Kolk and colleagues (2003) that the P600 effect is a reflection of the control operation triggered by a mismatch between the thematic interpretations proposed by the semantic heuristic and the syntactic parser. To this aim, we used the same stimulus materials as in the study of Kolk et al. (2003). However, there was one essential difference in the way participants were instructed in the present experiment. In the Kolk and colleagues' study (2003), participants were asked to indicate if the sentence was semantically plausible or not. Implausible was defined as semantically unacceptable. In the present experiment, participants were told that semantic reversals had been created on purpose and that they should not be misled by their knowledge of what normally happens in the world, but pay extra attention to "who does what to whom" and to evaluate whether this scenario fits well with their world knowledge or not. Note that this focus-on-syntax instruction creates a context in which semantically odd sentences are expected.

As stated above, a P600 effect is assumed to be triggered by a conflict between the outcome of a lexical strategy with that of the parsing routine. From this, we predicted that if the P600 effect after semantic reversal anomalies reported by Kolk and colleagues (2003) reflects a control process to check for possible processing errors, the P600 effect in the present experiment should be reduced. Since our focus-on-syntax instruction creates a context wherein anomalous interpretations are expected, the discrepancy between the thematic interpretations proposed by the semantic heuristic and the syntactic parse should be diminished and therefore less readily qualified as a possible processing error. Hence, we predicted that our focus-on-syntax instruction should diminish the inclination to re-attend to a possible processing error which should be reflected in a decrease in error rates and a reduction or even elimination of the P600 effect to semantic reversal anomalies.

2 Method

2.1 Participants

There were 38 participants (mean age = 22 years; age range = 18 to 30; 29 females). All were native speakers of Dutch, had no reading disabilities, were right-handed and had normal or corrected-to-normal vision. Hand dominance was assessed with an abridged Dutch version of the Edinburgh Inventory (Oldfield, 1971). Sixteen participants reported the presence of left-handedness in their immediate family.

2.2 Materials

The semantic list consisted of 68 Dutch sentences with centrally embedded relative clauses. For each sentence, a subject relative (SR) and an object relative (OR) version, a plausible and an implausible version were created, yielding a total set of 272 sentences (see Table 1). Fourteen (out of 68) sentences in the object relative condition employed adjuncts rather than prepositional complements. The semantically anomalous sentences expressed scenarios conflicting with general world knowledge (e.g., <u>foxes</u> are not very likely to be hunting <u>poachers</u> whereas poachers are likely to hunt foxes). The anomalies resulted from reversing the first and the second noun phrase of semantically acceptable sentences. The two noun phrases could both serve as the agent and the patient of the action expressed by the verb ending the relative clause (e.g., <u>foxes</u> and <u>poachers</u> can hunt as well as be hunted). The anomaly was not evident before the relative clause's verb. This was done to ensure that the detection of the anomaly required deep processing of the relative clause, in that it depended on the successful integration of the verb with <u>both</u> noun phrases. In half of the sentences, the two noun phrases had the same grammatical number and in the other half they had a different number (singular or plural). The four versions of each sentence were counterbalanced across lists. Each list contained 17 SR acceptable sentences, 17 SR semantically anomalous sentences, 17 OR acceptable sentences, and 17 OR semantically anomalous sentences. Sixty-eight filler sentences were added to each list: 17 acceptable right-branching sentences, 17 semantically anomalous right-branching sentences (e.g., *De rechter luisterde naar de beklaagde die opkwam voor zijn advocaat*. Literal translation: The judge listened to the defendant who stood up for his lawyer.), 17 acceptable conjunctions and 17 conjunctions with a semantic reversal anomaly (e.g., *De zeehonden doken in het water en vingen de ijsbeer.* Literal translation: The seals plunged into the water and caught the polar bear.).

Table 1 Examples of the acceptable and unacceptable versions of the sentences separately for the 2 levels of complexity

	Plausible sentence	Implausible sentence
Subject relative	De docent die aan de studenten *lesgaf* kwam het lokaal in.	De studenten die aan de docent *lesgaven* kwamen het lokaal in.
Word-by-word translation	The teacher who on the students gave [single] lesson entered the room.	The students who on the teacher gave [plural] lesson entered the room.
Paraphrase	The teacher who taught [singular] the students entered the room.	The students who taught [plural] the teacher entered the room.
Object relative	De studenten aan wie de docent *lesgaf* kwamen het lokaal in.	De docent aan wie de studenten *lesgaven* kwam het lokaal binnen.
Word-by-word translation	The students to whom the teacher lesson gave [singular] entered the room.	The teacher to whom the students gave [plural] lesson entered the room.
Paraphrase	The students who were taught [singular] by the teacher entered the room.	The teacher who were taught [plural] by the students entered the room.

2.3 Procedure

Participants were seated in a closed chamber. A response device with three push-buttons was fixed on a small table in front of the participant. Sentences were presented in serial visual presentation mode at the center of a PC monitor. Word duration was 345 ms and the stimulus-onset asynchrony (SOA) was 645 ms. Sentence final words were followed by a full stop. The inter-trial interval was 2 seconds. Words were presented in black capitals on a white background in a 9 cm by 2 cm window at a viewing distance of approximately 1 m. Each sentence was preceded by a fixation cross (duration 510 ms) followed by a 500 ms blank screen.

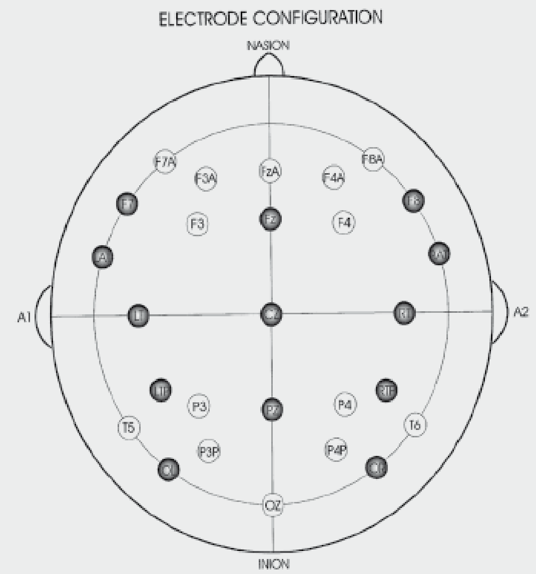
Participants were told that semantic anomalies had been created on purpose, by reversing the agent and patient of otherwise normal sentences (fox is hunting poacher instead of poacher is hunting fox). They were asked to attend carefully to the structure of the sentences and evaluate who does what to whom. So, participants had to pay close attention to who was the agent and who was the patient of the sentence and indicate whether this fit well with their world knowledge or not. The experimenter pointed out that it was important that the participants should not be misled by their knowledge of what normally happens in the world (that is, normally poachers are hunting foxes and foxes are being hunted) but attend to the structure of the sentences. Incongruent was defined as unlikely based on our knowledge of the world. During a short training the detection of semantic incongruencies was practiced with 10 sentences on paper. The experimenter turned the participants' attention to the reversals in each of the 10 sentences and explained how they were constructed. Participants were subsequently instructed to attentively read each sentence presented on the computer screen and to press a button with the dominant index finger if the sentence was congruent with world knowledge and with the other index finger if it was not. See Table 1 for the example sentences.

The list was split up into 5 blocks; there was a brief pause between blocks and each block was preceded by two filler items. Participants had to postpone their acceptability judgment until presentation of a prompt that occurred 1500 ms after the sentence final word. We used a delayed response task to eliminate effects of motor response preparation on the ERPs of interest. Because eye movements distort the EEG recording, participants were trained to make eye movements, e.g. blinks, only in the period that the prompt was present (stimulus duration was 2295 ms). Prompt offset was followed after 705 ms by a fixation cross indicating the start of the next trial.

2.4 EEG data acquisition and analysis

EEG was recorded with 27 tin electrodes mounted in an elastic electrode cap (Electrocap International). Figure 1 presents the electrode configuration.

Figure 1 Electrode configuration used in the present experiment. The electrode configuration of the Kolk et al. (2003) study is displayed by the bold circles



The electrode positions included standard International 10-20 system locations over the left and right hemispheres at the frontal (F3, F4, F7 and F8), midline (Fz, Cz, Pz, Oz), parietal (P3, P4) and temporal (T5, T6) sites. Eight extra electrodes were placed at the frontal (F3A, FZA, F4A, F7A, F8A), midline (OZ) and parietal (P3P, P4P) sites. In addition, eight electrodes were placed at non-standard electrode positions previously found to be sensitive to language manipulations (e.g., Holcomb & Neville, 1990): left and right anteriortemporal sites (LAT and RAT: 50% of the distance between T3/4 and F7/8), left and right temporal sites (LT and RT: 33% of the interaural distance lateral to Cz), left and right temporoparietal (LTP and RTP: Wernicke's area and its right hemisphere homologue: 30% of the interaural distance lateral to a point 13% of the nasion-inion distance posterior to Cz), and left and right occipital sites (OL and OR: 50% of the distance between T5/6 and O1/2). The left mastoid served as reference. Electrode impedance was less than 3 KΩ. The electro-oculogram (EOG) was recorded bipolarly; vertical EOG was recorded by placing an electrode above and below the right eye and the horizontal EOG was recorded via a right to left canthal montage. The signals were amplified (time constant = 8 s, bandpass = 0.02 – 30 Hz), and digitized online at 200 Hz. Presentation of stimuli and recording of performance data was accomplished by a Macintosh computer.

EEG and EOG recordings were examined for artifacts and for excessive EOG amplitude (>100 μV) extending from 100 ms before the onset of the critical verb ending the relative clause to 1000 ms following its onset. Averages were aligned to a 100-ms baseline period preceding the critical verb. Based on previous studies using the same or similar materials (Kolk et al. 2003; Van Herten et al. 2005), mean amplitudes were calculated in the time windows of 400-500 ms and 650-850 ms to capture N400 and the P600 effects, respectively. Another reason these windows were chosen was to make a direct comparison between our dataset and the dataset of Kolk et al. (2003) possible. For both time-windows, separately for the midline and the lateral sites, repeated measures MANOVAs were conducted with plausibility (plausible vs. implausible) and complexity (SR vs. OR) as within subject factors. For the midline sites the additional factor was site (Fza, Fz, Cz, Pz, Oz). To further explore the scalp distribution of the ERP effects for the lateral sites we used a region of interest (ROI: anterior vs. posterior) by hemisphere by lateral site (F7a/F3a/F7/F3 vs. P3/P3p/T5/OL vs. F8a/F4a/F8/F4 vs. P4/P4p/T6/OR) design. Relevant interactions with site and plausibility were followed up by post hoc Newman-Keuls' tests to assess the significance of contrasts. The multivariate approach to repeated measurements was used to avoid problems concerning sphericity (e.g., Dien & Santuzzi, 2004; Vasey & Thayer, 1987).

3 Results

3.1 Reaction time and error data

The reaction time (RT) and error data were entered into separate repeated measures multivariate analyses of variance (MANOVAs) with plausibility (plausible vs. implausible), and complexity (SR vs. OR) as within subject factors. Mean RT and error percentages are presented in Table 2.

Table 2 Mean reaction time (RT) and error percentages (Error) with standard deviations (S.D.), for the plausible, implausible and subject-relative, and object-relative sentences

	Subject relative				Object relative				RT		Error	
	RT	S.D.	Error	S.D.	RT	S.D.	Error	S.D.	Mean	S.D.	Mean	S.D.
Plausible	532	29	6.52	.01	587	25	12.08	.02	560	25	9.30	.01
Implausible	502	25	3.97	.01	552	27	8.11	.01	527	24	6.04	.01
Mean	517	26	5.25	.01	570	23	10.1	.01				

The means are marginal means averaged over either complexity or plausibility.

The RT analysis revealed main effects of complexity, $F(1,36) = 21.32$, $p < .001$, and plausibility, $F(1,36) = 5.86$, $p = .021$. As Table 2 shows, the complexity effect indicated that mean RT for OR sentences (570 ms) was longer than mean RT for SR sentences (517 ms). The plausibility effect indicated that mean RT for plausible sentences (560 ms) was longer than mean RT for implausible sentences (527 ms). There was no complexity by plausibility interaction, $F < 1$.

The error analysis revealed main effects of complexity, $F(1,36) = 19.21$, $p < .001$, and plausibility, $F(1,36) = 8.9$, $p = .005$. These effects indicated that participants made more errors on OR (10%) than on SR sentences (5%) and that participants made more errors on plausible (9%) than implausible sentences (6%) [4]. No complexity by plausibility interaction was observed, $F < 1$.

Compared to the Kolk et al. (2003) study, participants in the present study appeared to have faster reaction times and lower error percentages. To test the significance of this difference we entered both the RT and error data of the Kolk et al. (2003) and the present study into a separate repeated measures MANOVA, with instruction (semantic plausibility judgment vs. focus-on-syntax) as between-subject factor and complexity (SR vs. OR) and plausibility (plausible vs. implausible) as within-subject factors.

3.1.1 Global analyses

The RT analysis revealed a main effect of instruction, $F(1, 76) = 33.79$, $p < .001$. Mean RT was faster in the present study (543 ms) than in the Kolk et al. (2003) study (861 ms). No interactions were obtained, $Fs < 2$. The error analysis indicated that participants in the present study were more accurate than those in the Kolk et al. (2003) study, 8% vs. 11% of errors respectively, $F(1, 76) = 5.32$, $p = .024$, indicating the there was no speed accuracy trade off. No further interactions were observed, $Fs < 3.5$.

[4] As expected, participants processed the OR sentences more slowly than the SR sentences, in addition they were less accurate on the OR sentences. This pattern is consistent with the more complex syntactic structure of OR sentences. On the other hand, plausible sentences were responded to more slowly than implausible ones and elicited more errors. This pattern appears unexpected since one might expect plausible sentences to be processed faster and/ or more accurately, but this pattern was also observed by Kolk et al. (2003). They explained it in the following way. Because participants had to wait until they had read the last word of the sentence before they could know that the sentence was plausible, they had to postpone their answer until the last word of the sentence was presented. However, participants could know that a sentence was implausible as soon as the anomalous verb was read. It is possible, therefore, that in the case of an implausible sentence, the decision could be made earlier in the sentence and participants could already prepare their response during the sentence. Perhaps, this led to faster mean reaction times and lower error percentages for the implausible sentences..

The global analyses thus confirmed that the focus-on-syntax instruction affected the behavior of the participants in the predicted direction. Participants were less easily misled by the semantic reversal anomalies when they were previously informed about the presence of semantic reversal anomalies and were instructed to attend to the syntax. This is reflected by the fact that participants in the present study were faster and more accurate than those in the Kolk et al. (2003) study, in which participants' attention was not directed to the sentence structure.

3.2 Event-related potentials

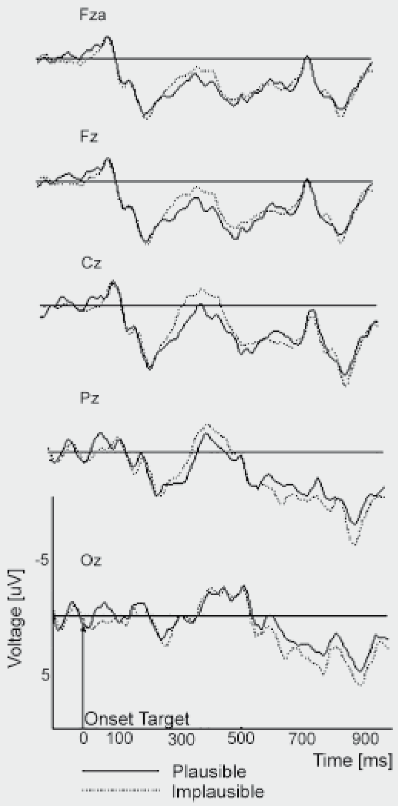
The grand mean ERPs to the critical verbs are presented in Figure 2 for the midline sites and in Figure 3 for the lateral sites. As can be seen in these Figures, the critical verbs elicited a negativity peaking at about 200 ms (N1) and a positivity peaking around 250 ms after the critical verb (P2) which were both maximal at the occipital sites. These components were followed by a broad negative wave in the 250-500 ms epoch peaking at about 350 ms, the N400, which was largest at central and posterior sites. It is well known that the N400 component is elicited by each open class word (e.g., Kutas & Van Petten, 1994). Inspection of the waveforms suggested that mean amplitude was more negative for implausible verbs than for plausible verbs at some midline sites (Fza, Fz, Cz, and Pz) and right hemisphere sites (RAT, RT, RTP).

The N400 was followed by a slow positive shift, the P600, starting at about 600 ms and extending up to 1000 ms which was largest at central and posterior sites. Inspections of the waveforms suggests the presence of a small P600 effect at the midline (see for example Pz and Oz) and at some lateral sites of the right hemisphere (see for example P4P or OR). The differences between conditions were, however, rather small (about 1 μV or less).

3.3. Statistical analyses

About 11% of the trials were excluded from the analyses because of artifacts; of which 2% belonged to the SR-plausible condition, 2% belonged to the SR-implausible condition, 4% belonged to the OR-plausible condition, and 3% belonged to the OR-implausible condition.

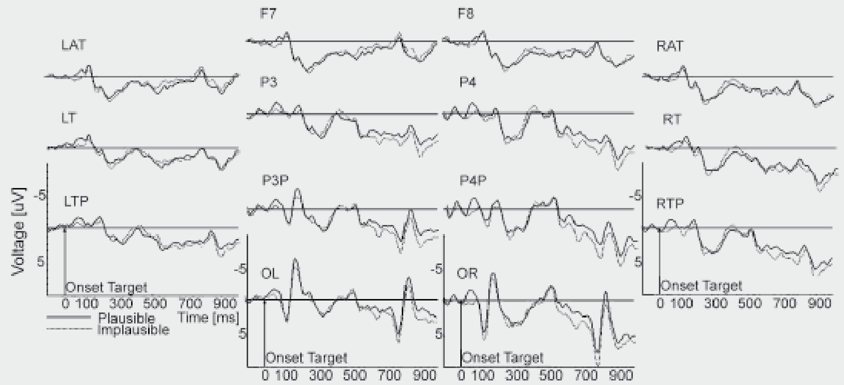
Figure 2 Grand ERP averages to the critical verb for the midline sites, collapsed over the two levels of complexity. Averages are time locked to the onset of the critical verb, and superimposed for the two levels of Plausibility. Negativity is plotted upwards.



3.3.1. N400 window (400 - 500 ms)

For the midline sites, the analyses for the N400 window did not reveal a main effect of plausibility, F < 3. No main effect of complexity, F < 1 or two- way interactions between plausibility and complexity were obtained, Fs < 1. No other interactions with plausibility were observed, all Fs < 2. For the lateral sites, no plausibility effects were obtained, F < 1. No main effect of complexity or interactions with plausibility and

Figure 3 Grand ERP averages to the critical verb for a representative subset of lateral sites, collapsed over the two levels of complexity. Averages are time locked to the onset of the critical verb, and superimposed for the two levels of Plausibility. Negativity is plotted upwards.



complexity were obtained, F < 1. In addition, no interactions of plausibility with site, ROI and/or hemisphere were obtained, all Fs < 3.5. Additional MANOVAs for the midline and the lateral sites in which N400 was measured in a broader latency window 300 – 500 ms following verb onset) revealed essentially the same pattern of results, in that no effect of plausibility or relevant interactions were obtained, all Fs < 3. [5]

In sum, the N400 analyses indicated that no N400 effect was observed, neither at the midline nor at the lateral sites. Therefore, no evidence for an N400 plausibility effect was obtained in the present experiment.

[5] ERP grand mean inspection suggests that an N400 is present in the 250-450 ms. interval for some electrodes of the midline and the right hemisphere. However, statistical analyses in this latency window reveal that these effects are not reliable. We decided to present the results for the 400-500 ms. interval in the main text, to be able to directly compare our data set with the data set of Kolk et al. (2003).

3.3.2. P600 window (650 - 850 ms)

For the midline sites no effect of plausibility, $F < 1$ or effect of complexity was observed, $F < 2$. In addition, no plausibility by site or other interactions with plausibility or complexity were obtained, all $Fs < 1$. In other words, there were no indications for a P600 effect at the midline sites. For the lateral sites, no main effect of plausibility was obtained, $F < 2$. However, an interaction of plausibility by ROI was found, $F (1, 35) = 5.89$, $p < .03$. Separate analyses for the two regions of interests indicated that a P600 effect was present at posterior sites, $F (1,35) = 6.41$, $p < .02$, but not at anterior sites, $F < 1$. Moreover, a trend for a four-way interaction between plausibility, ROI, hemisphere and site was found, $F (3, 33) = 2.54$, $p < .08$. To further examine the scalp distribution of the P600 effect post hoc Newman-Keuls' tests were conducted. These tests indicated that a P600 effect was present at the following four sites of the right hemisphere (P4, T6, P4P, OR: $p < .05$) and one single site of the left hemisphere (P3P, $p < .05$). [6]

In sum, the analyses for the P600 for the midline sites demonstrated that at centroparietal sites (Cz and Pz) which are the sites that typically show largest P600 effects to syntactic anomalies, no P600 effect was present. However, the analyses for the lateral sites indicated that the P600 effect was not totally eliminated; since a P600 effect was still obtained at some posterior sites. Thus, although in the present study there was no P600 effect for the midline sites, the latter (lateral) analyses suggest that subjects do not have complete control over their processing strategies.

The results for the N400 are similar to those of the study of Kolk et al. (2003), in that no reliable N400 effect was found. The results for the P600 on the other hand, differ from those of Kolk et al. (2003), because in that study clear P600 effects were obtained, both

[6] ERP grand mean inspection suggests that a P600 is present in the 550-750 ms. interval for some midline and posterior electrodes. However, statistical analyses in this latency window reveal essentially the same pattern of results as was revealed in the 650-850 ms. window. We decided to present the results for the 650-850 ms. interval in the main text, to be able to directly compare our data set with the data set of Kolk et al. (2003).

for the midline and for the lateral sites of the left and right hemisphere. The P600 effect in the present study, therefore, was mainly limited to right posterior sites.

3.3.3. Global analyses.

To directly compare the P600 effects, supplementary global analyses for P600, with instruction (semantic plausibility judgment vs. focus-on-syntax) as between-subject factor and plausibility (plausible vs. implausible) as within-subject factor were carried out. The relevant question was whether a plausibility by instruction interaction would be obtained. The main results of these global analyses were as follows: neither for the midline nor for the lateral sites an effect of instruction and/or an instruction by plausibility interaction was present, $Fs < 1.5$. Furthermore, no interaction of instruction with site and/or hemisphere was obtained that pointed to a difference in P600 pattern between experiments.

One could argue that given that the RTs in the present study were faster than those in the Kolk et al. study, that the use of the same broad window to quantify P600 effects might not be the best window to capture P600 differences between the two studies. [7] Therefore, supplementary analyses for the P600 peak amplitude, were carried out across studies. To this aim, the P600 peak amplitude was measured between 500 ms and 900 ms after word onset. The factors of the global analyses were: instruction (semantic plausibility judgment vs. focus-on-syntax) as between-subject factor and plausibility (plausible vs. implausible) as within-subject factor. The main results of the analyses for P600 peak amplitude were as follows: for the lateral sites, no effect of instruction and/or an instruction by plausibility interaction was present, $Fs < 1.5$.

[7] To examine more closely the onsets and lengths of the differences in the ERP plausibility effect between the two studies, supplementary time course analyses were conducted both for the 300 to 500 ms window (to capture N400) and the 500 to 900 ms window (to capture P600) using consecutive bins of 50 milliseconds. To reduce the chance of Type I errors due to the large number of comparisons, an effect is referred to as significant only if it was present in at least two consecutive time epochs. For both kinds of analyses the mean amplitudes were entered into a MANOVA. These analyses showed that neither for the N400 nor for the P600 earlier or more transient reliable differences between the two studies were present.

Furthermore, no interaction of instruction with site and/or hemisphere was obtained that pointed at differences in P600 pattern for the lateral sites between experiments.

However, most important for our present purposes, the analyses for the peak amplitude for the midline sites confirmed that there were indeed differences in P600 pattern between the two experiments. For the midline sites, the interaction between plausibility and instruction was significant ($F(1, 74) = 5.37$, $p < .05$). Separate analyses for the two experiments revealed that a P600 effect was present in the study of Kolk et al., $F(1,39) = 6.46$, $p < .01$, but not in the present study, $F < 1.5$.

Taken together the results from the present experiment and those from the global analyses support the following findings: the focus-on-syntax instruction was successful in reducing the P600 effect. This was reflected by: First, the fact that no P600 effect was present at the midline. The global analyses confirmed a change in P600 pattern as a function of instruction: In particular, the analyses for the peak amplitude showed that a P600 effect was obtained for the midline sites in the study of Kolk et al. (2003) but was absent in the present experiment. Second, opposite to the Kolk et al. (2003) data, the analyses for the lateral sites showed that the P600 effect in the present study was not bilaterally distributed, but mainly limited to a subset of posterior sites over the right hemisphere. Thus, the P600 effect did not totally disappear, as it was still present for a small number of posterior sites. In other words, the change in instruction led to a significant decrease in P600 effect. However, it did not completely abolish the P600 effect.

4 Discussion

It was argued in the Introduction that the language comprehension system uses a combination of heuristics and algorithms (Townsend & Bever, 2001; Fereirra, 2003). Usually, these heuristics and syntactic algorithms produce the same thematic interpretations. Kolk and colleagues (2003) studied sentences in which these two routes produced a conflicting result; semantic reversal anomalies were used which were formed by exchanging the subject and object of semantically acceptable sentences. They found a P600 effect, and not an N400 effect in response to these semantic reversal anomalies. Kolk and colleagues (2003) argued that the two routes produce

a mismatch between the semantically plausible thematic interpretation proposed by the semantic heuristic which is highly expected based on our world knowledge and the implausible one proposed by the algorithmic parser. This mismatch brings the brain to re-attend the unexpected element to check upon its veridicality. Kolk and colleagues (2003) proposed that this control operation underlies the P600 effect to semantic reversal anomalies. Since the word-level integration process does not encounter difficulties when integrating the words of a reversal anomaly into a coherent meaning, no N400 effect was obtained.

The present study tested the hypothesis that the P600 effect to semantic reversal anomalies is a reflection of the control operation triggered by the mismatch between thematic interpretations. The focus-on-syntax instruction created a context for participants wherein anomalous sentences were expected; they were explicitly informed that semantic reversals had been constructed on purpose and that they should not be misled by word meanings (‘knowledge of what normally happens in the world’), but attend to the structure of the sentences (‘who does what to whom?’). This instruction should reduce the discrepancy between the thematic interpretations proposed by the semantic heuristic and the syntactic parse. Consequently, semantic reversals were expected and therefore less readily qualified as a possible processing error. So, the necessity of the brain to re-attend the event to check for a possible processing error should be reduced. If the proposal by Kolk et al. (2003) is right, then our focus-on-syntax instruction should lead to a reduction or disappearance of the P600 effect to semantic reversal anomalies.

The major result of the present article is that our focus-on-syntax instruction did influence both the behavioral data and the ERP data. First, participants in the present study were faster and more accurate than the participants in the Kolk and colleagues’ (2003) study. This improvement in performance is taken to indicate that the instruction to focus on syntax and not on word meaning was effective, in that our participants were less easily misled by the semantic reversals.

Let us now turn to the ERP effects. The analyses for the midline sites revealed that there were no indications for a P600 effect at the midline sites; which are the sites that typically show the largest P600 effects to syntactic violations. In addition, this result stands in sharp contrast to that of Kolk et al. (2003), who reported a plausibility effect for the midline sites. The global analyses for P600 amplitudes further supported this difference in P600 pattern as a function of instruction. Importantly, an interaction between plausibility and instruction was obtained for the midline sites. Follow up tests verified that this interaction reflected the presence of a P600 effect in the Kolk et al. study ($p < .05$) but absence of this effect in the present study ($F < 1.5$). In addition, at lateral sites the P600 in the present study was reduced in that it was less broadly distributed than the typical syntactic P600 effect. A two-way interaction of plausibility by ROI revealed that the P600 effect was not totally eliminated in the present study, but that a P600 effect was still present for a small set of posterior sites over the right hemisphere. This also stands in contrast to the results of Kolk and colleagues (2003), who reported a main effect of plausibility for the lateral sites of both the left and right hemisphere.

The focus-on-syntax instruction thus led to a disappearance of the P600 effect at the midline sites and at all but one site of the left hemisphere. We propose that this instruction directed the participants' attention to the structure of the sentences and created a context wherein semantic reversals were expected. The mismatch in the thematic interpretation proposed by the heuristic and the one proposed by the parser was expected and therefore not qualified as a likely processing error. Because the instruction turned an unexpected real life event (that foxes hunt poachers) into a less unexpected event, there was less need for the brain to re-attend the implausible linguistic unit which resulted in a decrement of the P600 effect. Hence, the results of the present study support the proposal by Kolk and colleagues (2003) that the P600 effect to semantic reversal anomalies is based on a control operation triggered by a mismatch in thematic interpretations.

The fact that a residual P600 effect was still obtained for a small set of posterior sites suggests that the reversal anomalies did to some extent still elicit a mismatch in the thematic interpretation proposed by the semantic heuristic and the one proposed by the syntactic algorithm. Apparently, participants did not have complete control over their natural tendency to give priority to semantic processes. In spite of the focus-on-syntax instruction, the bias for the semantically most plausible interpretation continued to be active, though to a lesser degree. What is critical for the current purposes is that this P600 effect was less broadly distributed than the P600 effect observed in the Kolk and colleagues' study (2003), where the same semantic

reversals had been presented. This suggests that a substantial part of the P600 effect after semantic reversals can be accounted for by the control operation that is triggered by a mismatch between two thematic interpretations.

The observed reduction in P600 effect is consistent with previous studies that have shown that the P600 effect is affected by task demands and is therefore assumed to reflect a process that is largely under the participant's control (Gunter & Friederici, 1999; Gunter, Stowe, & Mulder, 1997; Hahne & Friederici, 1997). For example, the amplitude of the P600 has been shown to be modulated by probability with larger P600 effects to the less probable event (Coulson, King, & Kutas, 1998; Hahne & Friederici, 1999; Gunter et al., 1997). This indicates that the P600 effect is sensitive to list composition that affects subjects' expectations.

As shown by previous studies, the P600 effect to syntactic and certain semantic violations shows a central/posterior scalp distribution (Coulson et al., 1998; Kolk et al., 2003). The residual P600 in the present experiment (after a focus-on-syntax instruction) had a slightly more posterior scalp distribution, including the right occipital site. Note that also in the Kolk et al. study significant P600 effects were present at occipital sites. At least for the data presented by Kolk et al. (2003), which allowed a direct within- subject comparison between the topography of the P600 after semantic reversal anomalies and the syntactic violations, there were no indications for topographical differences in P600 effect. In the present study, such a within-subject comparison is impossible. Based on the Kolk et al. study, though, we consider the possibility that the present (semantic) P600 effect is qualitatively different in terms of scalp distribution as rather unlikely. Furthermore, we would like to point out that there is evidence for some variation in scalp distribution of the syntactic P600 effect. A more frontal/broad distribution of the P600 effect has for example been reported for locally ambiguous sentences (Friederici et al., 1996; Hagoort et al., 1999; Osterhout and Holcomb, 1992; Van Berkum et al., 1999) while Kaan and Swaab (2003) observed a more posterior distribution.

Bever and colleagues (1998) propose that the semantic interpretation developed by the heuristic depends on passively accumulated and applied statistical generalizations. Thus, according to this proposal, the initial semantic analysis proceeds on the basis of statistically sensitive perceptual strategies. As described in the Introduction, the most strongly confirmed abstract pattern available to the heuristic is the word-order strategy. This implies that a noun phrase preceding and agreeing in number with a verb is taken as the subject of that verb. Similarly, a noun phrase following a verb is taken to be the object of the verb. However, this proposal was not supported in the present study. If participants would assign this 'favoured' canonical form to our semantic reversal sentences, they would not have been misled by our reversals. In a sentence like "The fox that <u>hunted</u> [singular] the poachers stalked through the woods", the noun phrase preceding the verb is the subject of the verb, and the noun phrase following the verb is the object of the verb. But, the present experiment indicated that participants were, though to a lesser degree compared to the Kolk and colleagues study (2003), misled by the reversals. This is bolstered by the error percentages (participants had higher error percentages for the implausible sentences compared to the plausible sentences) and the residual P600 effects for the implausible compared to the plausible sentences.

No evidence was obtained for an N400 effect in the present study. Consistent with our previous reasoning, this is because the content words of the semantic reversals in the implausible and plausible sentences were comprised of the same lexical items that were easy to integrate into a coherent meaning. No integration difficulty and hence no N400 effect was expected to occur. This supports the claim that the semantic heuristic develops a thematic interpretation on the basis of the *meaning* of the individual words. So, participants are not biased to follow the canonical form, but they are biased by the meanings of the individual words and their world knowledge. This is in agreement with Bever's (1970) description of 'Strategy C' and Ferreira's (2003) description of the 'plausibility heuristic'. These heuristics propose a semantic interpretation of an utterance which is in agreement with knowledge of the individual content words or schema's in long term-memory. The use of heuristics has also been shown to play an important role in the explanation of comprehension difficulty by aphasic patients (Saffran et al., 1989).

As presented in the Introduction, 'syntax-first' models propose that a modular syntactic processing system guides and precedes semantic interpretation. Semantic information is used only in a second stage, either to choose between different structural possibilities or to guide revision after the chosen structure turned out to be erroneous. The syntactic structure of our semantic reversals is unambiguous.

That is, in the reversal sentence the fox that hunted the poacher, there is only one option: fox is the Agent, and poacher is the Theme. A syntax-first model would thus predict a semantically incorrect representation of this sentence. This would predict the modulation of the N400 which, however, was not found. Constraint-based models propose that semantic information is used during syntactic structure build-up. Hence, semantic information may help to choose between different structural possibilities. But, semantic information cannot propose structural possibilities. If syntactic cues are unambiguous, then semantic information does not exert a controlling influence. This implies that the constraint-based models would also predict an N400 effect after our semantic reversals. Hence, both constraint-based models as syntax-first models assume that syntactic information, when unambiguous, will control the initial combinatory analysis of linguistic input. Contrary to what both models predict, we report that semantic information can independently propose a thematic interpretation of a sentence, overwhelming unambiguous syntactic cues. The data from our laboratory seem to be more consistent with a language processing system of parallel, independent syntactic and semantic processing mechanisms; in which semantic processing can overrule unambiguous syntactic cues (see also, Kim & Osterhout, 2005).

In conclusion, our results support the idea that the P600 effect to semantic reversals is based on a control operation after a mismatch in thematic interpretations to check upon the possibility of a processing error. The present results show that the focus-on-syntax instruction recruited the neural system underlying the control operation to a lesser degree than the semantic plausibility judgment instruction in the experiment by Kolk and colleagues (2003). This is demonstrated by the disappearance of the P600 effect most notably at central and parietal midline sites (which typically show maximal P600 effects) and the fact that the P600 effect was not bilaterally distributed but mainly limited to the posterior areas over the right hemisphere. However, the residual P600 effects indicate that the focus-on-syntax instruction did not completely resolve the conflict between the thematic interpretations proposed by the semantic heuristic and the syntactic parse as a possible reading error. This suggests that the semantic heuristic and hence the control operation to check for a processing error continued to be active to some extent. In the face of strong expectations which are incongruent with the syntactically driven interpretations, it may be difficult to block the monitoring response completely.

Acknowledgments

Portions of this research were presented at the 11th Annual Meeting of Cognitive Neuroscience Society in San Francisco, 2004. We thank two anonymous reviewers for their very helpful comments on an earlier version of this article. We are grateful to the ERG group of the NICI for technical assistance and to Maarten vd Meulen for his support.

References

Besson, M., Fischler, I., Boaz, T., & Raney, G. (1992). Effects of automatic associative activation on explicit and implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 89-105.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362). New York: Wiley.

Bever, T.G., Sanz, M., & Townsend, D.J. (1998). The emperor's psycholinguistics. *Journal of Psycholinguistic Research*, 27, 261-284.

Chomsky, N. (1957). *Syntactic structures*. The Hague-Paris: Mouton & Co.

Chwilla, D.J., Brown, C.M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology,* 32(3), 274-285.

Chwilla, D.J., Hagoort, P., & Brown, C.M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51A(3), 531-560.

Chwilla, D.J., Kolk, H.H.J., & Mulder, G. (2000), Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, 42(3), 314-341.

Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: event-related brain response to morphosyntactic violations. *Language and cognitive processes*, 13(1), 21-58.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing : A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.

Dien, J., & Santuzzi, A.M. (2004). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In: T.C. Handy (Ed.), *Event-Related potentials: A methods book* (pp. 57-82). Cambridge, Massachusetts: The MIT press.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47,164-203.

Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of memory and language*, 25, 348-368.

Fodor, J.A. (1983). *The Modularity of Mind.* Cambridge, MA: Bradford Books.

Frazier, L., & Fodor, J.D. (1978). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 291-325.

Friederici, A.D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1219-1248.

Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103, 650-669.

Gigerenzer, G., Todd, P.M., & ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gunter, T.C., & Friederici, A.D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, 36, 126-137.

Gunter, T.C., Stowe, L.A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660-676.

Hahne A., & Friederici, A.D. (1997). Two stages in parsing: Early automatic and late controlled processes. *Experimental Brain Research*, 117, S47.

Hahne, A., & Friederici, A.D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194-205.

Hagoort, P., Brown, C.M., & Osterhout, L. (1999) The neurocognition of syntactic processing. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273-317). Oxford: University Press.

Hoeks, J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading, *Cognitive Brain Research*, 19, 59-73.

Holcomb, P.J., & Neville, H.J. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes*, 5, 281-312.

Kaan, E., & Swaab, T.Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15, 98-110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: a study with event related potentials. *Brain and language*, 85, 1-36.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, 21, 489-530

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117-129.

Kutas, M., (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533-572.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463-470.

Kutas, M., & Hillyard, S.A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience,* 1, 38-49.

Kutas, M., & Schmitt, B.M. (2003). Language in microvolts. In M.T. Banich, & M. Mack (Ed), *Mind, brain, and language: Multidisciplinary perspectives* (pp. 171-209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kutas, M., & Van Petten, C.K.F. (1994). Psycholinguistics electrified. Event-related brain potential investigations. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-143). San Diego, CA: Academic Press.

McDonald, M.C., Maryellen, C., Pearlmutter, N.J. & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97-113.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Saffran, E.M., Schwartz, M.F., & Linebarger, M.C. (1989). Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language,* 62, 255-297.

St John, M.F., & McClelland, J.L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46, 217-257.

Townsend, D.J., & Bever, T.G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: The MIT Press.

Trueswell, J.C., & Tanenhaus, M.K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, K. Rayner & L. Frazier (Eds.), *Perspectives on sentence processing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Vasey, M.W., & Tayer, J.F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution, *Psychophysiology*, 24, 479-486.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Chapter 3

Monitoring in language perception: The effect of misspellings of words in highly constrained sentences[1]

[1] This chapter has been published as Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Abstract

We present evidence for a monitoring process in language perception at the word level, reflected by a P600. This P600 is triggered when a conflict evolves because the brain encounters an unexpected linguistic item when another item is highly expected. To resolve this conflict between representations, the brain monitors the input to check for possible processing errors. A P600 was hypothesized to occur after orthographic anomalies, like pseudohomophones, in particular when the word from which the pseudohomophone is derived is highly expected. This hypothesis was tested by recording ERPs while participants read high-cloze sentences ('In that library the pupils borrow *books* ....') and low-cloze sentences ('The pillows are stuffed with *books* ....'). In a pretest, the high-cloze sentences were produced by more than 90% of the subjects, while the low-cloze sentences were never produced. In half of the sentences the critical word *books* was replaced by a pseudohomophone (e.g., *bouks*), which in the high-cloze sentences orthographically and phonologically resembles the highly expected word. Consistent with the monitoring hypothesis, only pseudohomophones in high-cloze sentences elicited a widely distributed P600 effect while pseudohomophones in low-cloze sentences did not. A standard N400 effect of cloze probability occurred both for words and pseudohomophones. The present ERP results support the view that there is a process of monitoring that takes place in language perception which is reflected by the P600. It occurs whenever a conflict between a strong tendency to accept and one to reject a word brings the cognitive system in state of indecision.

Introduction

Monitoring refers to a process that evaluates the appropriateness or correctness of ongoing motor activity or response output. It is a function of cognitive control aimed at output optimalization: to bring erroneous behavior in line with desired goals (e.g. Botvinick et al., 2001; Postma, 2000).

In the language domain, monitoring can manifest itself in the phenomenon of *self-repair* in speech. In 'overt' self-repairs, speech is interrupted and a new attempt is made at producing the correct form (e.g., *'I thought she …I thought he was looking at me*). Levelt (1983) argues that in addition to overt repairs, there are also 'covert' repairs in which errors are intercepted at the level of planning by an inner monitoring mechanism. This inner monitoring mechanism hence operates on a prearticulatory representation of the utterance (pre-articulatory editing). Covert repairs manifest themselves as various speech disfluencies such as prolongations or pauses (*I thought… I thought he was looking at me*). An important argument for the existence of pre-articulatory editing is that sometimes repair occurs after just one phoneme has been produced. Considering the early moment of these repairs, it seems improbable that errors are always detected by the speaker while listening to her own overt speech. The dominant theory of error monitoring in speech is that the pre-articulatory and post-articulatory editing in speech is accomplished by the language comprehension system (Levelt, 1983). In other words, we use the same mechanism for comprehending speech and for monitoring our own speech. Hence, Levelt's "perceptual loop theory" localizes monitoring in the perception apparatus. Recently, Hartsuiker and Kolk (2001) have provided computational evidence for this theory.

In the action domain, studies of event-related brain potentials (ERPs) have revealed a brain response following errors: the error-related negativity (ERN), typically occurring around 100 ms after an error (for a review see Yeung, Botvinick, & Cohen, 2004). ERN activity is not only observed after errors in choice reaction time tasks but also when participants are told that an error occurred, whether this was true or not. If an overt error is not necessary for an ERN to occur, what is it that elicits the ERN? Recent theories suggest that it is the conflict between two representations that triggers the ERN. Recent evidence provides support for the hypothesis that this monitoring process is generated in the anterior cingulate cortex (ACC).

Monitoring has been studied primarily in production tasks. However, besides errors in production, we also make errors in perception (e.g. misreading a word) or comprehension (e.g., misunderstanding a speaker) and it seems likely that these errors are also monitored for. A monitoring process at the sentence perception level has been described by Kolk, Chwilla, Van Herten, and Oor (2003). They assume that in sentence perception, simple processing heuristics are used in addition to syntactic algorithms, which together determine the final interpretation of the sentence. Heuristics can be regarded as ‘rules of thumb’: highly economical strategies that are generally but not always effective in extracting meaning. The syntactic algorithm on the other hand involves an algorithmic analysis of the syntactic structure of the sentence; this analysis is time-consuming but always comes up with the correct sentence interpretation.

Ferreira (2003) has recently caught up on the discussion of the use of simple processing heuristics in language comprehension. She tested whether the participants’ performance on deciding on the thematic roles in sentences that varied in plausibility (plausible: the dog bit the man vs. implausible: the man bit the dog) and in reversibility (reversible: the dog bit the man vs. non-reversible: the mouse ate the cheese) could be modeled by the use of two simple heuristics. One heuristic is the NVN strategy; that is, the processor assumes that the subject is a proto-agent and the object is a proto-patient. The second heuristic is the plausibility heuristic which states that the processor assumes the semantic analysis which is most consistent with world knowledge. It thus combines lexical items of a sentence in the most plausible way (see also, Ferreira, Ferraro, & Bailey, 2002). Although the NVN strategy was the best predictor, the combined use of both strategies could mimic the participants’ performance even better. The study thus provides evidence for the use of both the NVN and the plausibility strategies in normal speakers.

As Ferreira points out, it is presently unknown how the product of the heuristics is coordinated with the output of the syntactic algorithms. Perhaps heuristics are employed when algorithms are hard to apply, given the complexity of the sentence. Alternatively, algorithms may be used when comprehenders have little confidence in the outcome of a heuristic. Although these possibilities exist, it seems simpler and more straightforward to assume that the two routes operate in parallel and largely independent from one another. Kolk et al (2003) and Van Herten, Chwilla,

and Kolk (2006) argue for the latter possibility and point to the dual route model of reading aloud, in which there is similar parallel processing along two independent routes (e.g., Coltheart, Curtis, Atkins, & Haller, 1993). But if there is such parallel processing in the case of sentence interpretation, it is possible that the two routes lead to conflicting outcomes. Kolk et al. (2003) propose that such a conflict triggers a monitoring process and that it is this monitoring process which underlies their ERP findings. In particular, they used semantic reversal anomalies, which were formed by exchanging the subject and object of semantically acceptable sentences such as (1).

(1) De vos die op de stropers joeg sloop door het bos (original).
The fox that at the poachers hunted [singular] stalked through the woods (literal translation).
The fox that hunted [singular] the poachers stalked through the woods (paraphrase).

It is clear that in these sentences plausibility heuristics and syntactic algorithms produce different thematic interpretations. Whereas the plausibility heuristic (a lexical strategy) leads to the interpretation that poachers are hunting foxes, the parsing routines lead to the interpretation that the foxes are hunting the poachers. Although the latter interpretation is not entirely impossible it represents a highly unlikely event based on world knowledge.

Kolk et al. (2003) observed a P600 effect to these semantic reversal anomalies, and not an N400 effect – as would have been expected, given that semantic anomalies typically elicit an N400 effect. This is consistent with recent findings from other researchers who, despite differences in sentence material and language (English and Dutch), observed a P600 effect in the absence of an N400 effect to semantically implausible sentences relative to their plausible counterparts (Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Kuperberg et al., 2006; Van Herten, Kolk, & Chwilla, 2005). These results seem to challenge the view that P600 effects are primarily elicited by syntactic anomalies. How does the occurrence of the P600 relate to the presence of a conflict between algorithmic and heuristic processing routes?

The P600 effect is assumed to reflect an immediate consequence of the situation that the parse and the plausibility heuristic suggest different interpretations. Kolk et al. (2003) argued that the language comprehension system attempts to resolve the resulting conflict by checking the input for possible processing errors. In particular, the mismatch between the semantically plausible, highly expected (based on world knowledge) thematic interpretation and the implausible thematic interpretation makes it necessary for the brain to re-attend the unexpected linguistic unit to check upon its veridicality. After all, an inconsistency can have two sources. It can be real, in the sense that an unexpected event has indeed occurred (e.g. man bites dog). On the other hand, it can also stem from a processing error. To prevent integration of erroneous information into the current discourse, the reader will generally monitor the correctness of his or her analysis in case of a conflict. This explains the occurrence of P600 effects to semantic reversal anomalies.

Kolk et al. (2003) proposed that the absence of an N400 effect was due to the fact that the 'lexical' interpretation for both the plausible and implausible sentences lead to an interpretation that is plausible. Since the plausibility heuristic does not have difficulty integrating the words of the semantic reversals into a coherent message (the lexical items in both conditions are the same; the fox and the hunters) readers initially do not notice the anomaly. Hence, no N400 effect was elicited. It thus seems that a *conflict* at the sentence level triggers the monitoring process in perception, just as in production. This conflict can probably best be described as one between different tendencies: the tendency to reject and the tendency to accept the sentence.

Thus, we propose that if language perception leads to the activation of two incompatible interpretations, a conflict would arise, signaling the possibility of a processing error. Such a conflict could trigger a monitoring response and hence a P600, to check for the possibility of a processing error. Garden path sentences are one example of a situation known to elicit P600 effects (Osterhout and Holcomb, 1992) which can be characterized as representing some kind of conflict between response tendencies. In garden path sentences, two different analyses of the same linguistic string lead to the activation of two incompatible responses. After reading the sentence 'The woman persuaded to answer the door', initially one interpretation is chosen, but has to be replaced by a different interpretation later on. At first,

readers assume that the sentence is about a woman persuading someone, but after reading the sentence part following the verb, they realize that the sentence is about a woman being persuaded. The brain resolves this indecisive state by monitoring, to check for the possibility of a processing error.

The goal of the present experiment was to test whether a monitoring process triggered by a conflict could also be present at the word level. In a study with German-speaking participants, Münte, Heinze, Matzke, Wieringa, and Johannes (1998) observed a P600 effect after orthographic anomalies (*Die Hexe benutzte ihren Behsen, um zum Wald zu fliegen*. Literal translation: The witch used her broome to fly to the forest.). Could this P600 effect stem from a conflict, similar to what we proposed for the semantic anomalies? One might indeed argue that there is a strong tendency to accept the word Behsen (broome). First of all, it is semantically highly expected. Secondly, the phonological form of the word confirms this expectation and makes it maximally strong. It is as if we ask a participant in a Cloze test to fill in a word that refers to something witches tend to use to fly to the forest, a word that sounds like /brum/. It would seem that this word is 100 % predictable [2]. On the other hand, there will be a strong tendency to reject the word since the orthographic form does not fit the phonological form: the word is misspelled. So we may indeed have a strong conflict between a tendency to accept and a tendency to reject the misspelled word in this context. To be sure, the source of the conflict is very different from what we saw in the case of the semantic anomalies. Here, the tendency to accept stemmed from the fact that a lexical strategy indicated a highly plausible interpretation and the regular parse a highly implausible interpretation. In both cases, a conflict seems to exist nevertheless. An interpretation cannot be simultaneously plausible and implausible. Similarly, a

[2] The dominant view in psycholinguistics is that language processing proceeds in a strict bottom-up fashion. However, there are at least some recent ERP studies that provide clear evidence for top down influences; that is, language users online generate expectancies for up-coming words. Of special interest is a study by deLong, Urbach, and Kutas (2005), who show that expectancies manipulate ERP responses in a graded fashion. In particular, they show that participants do not only generate expectancies in high constrained sentences but even in less constrained sentences (i.c. for cloze probability less than .5)

particular word cannot simultaneously be there and not be there. A conflict like this would bring the cognitive system into a state of indecision. Thereby, it would trigger a monitoring response, involving reprocessing the critical linguistic string to detect and restore a possible processing error.

In our test of the monitoring hypothesis at the word level, we followed Münte et al. (1998) by presenting pseudohomophones in high-cloze contexts. What we added was a low-cloze condition. So we created both a high and a low-cloze context for a particular lexical item (see Table 1). There was no difference in high and low-cloze context sentences except for the critical item ("The pillows are stuffed with *feathers* which make them feel soft." versus "The pillows are stuffed with *books* which makes them feel hard"). The critical lexical item was either spelled correctly or was a pseudohomophone derived from the expected word and phonetically similar to the expected word. The pseudohomophone was created by changing the vowel of the second syllable (e.g., pseudohomophone derived from the word 'boeken' is 'boekun'); so, the changed vowels were unstressed and always in the second part of the word.

The predictions were as follows. First, for the high-cloze context we expected the system to anticipate that a particular word will occur, and then start up a monitoring process when a different stimulus that is phonologically identical and orthographically similar to the highly expected item is actually presented. We predicted that this monitoring process will elicit a P600. Second, for the low-cloze sentences, the lexical items from which the pseudohomophones are derived are not highly expected and thus should not elicit a conflict between the expected and actually presented lexical item. Consequently, no monitoring process and hence no P600 effect was expected to occur. Third, we predict a standard N400 effect of cloze probability when comparing correctly spelled words in the high and the low-cloze condition. This is a classical finding but one may nevertheless wonder why we do not predict a P600 effect in this condition as well, since the low-cloze words are unexpected. As argued above, the function of the monitoring process is to edit out possible processing errors. Such a process should not be triggered by every single unexpected linguistic unit. Informative statements are always somewhat unexpected but they should be normally integrated into the discourse information. Checking for possible processing errors too often would disrupt communication. We thus

assume that mildly unexpected units are integrated in the discourse but that only highly unexpected units can lead to a conflict which triggers a monitoring response. These predictions would be in accordance with a study by Gunter, Friederici, and Schriefers (2000) in which low-cloze nouns elicited a larger N400 than the high-cloze nouns; and in which a P600 component was found only after gender disagreement in high-cloze nouns.

Fourth, if as proposed by Newman and Connolly (2004) semantic integration (i.e., the N400) is reliant on the phonological and not the orthographic representation, then N400 amplitude for the pseudohomophones should also be larger in the low-cloze than in the high-cloze condition. Finally, an early negativity to orthographic mismatches has

Table 1 Examples of the high-cloze and low-cloze versions and the correct word and pseudohomophone versions of the critical sentences

	Correct	Pseudohomophone
High cloze	In die bibliotheek lenen scholieren *boeken* om mee naar huis te nemen	In die bibliotheek lenen scholieren *boekun* om mee naar huis te nemen
Word-by-word translation	In that library borrow the pupils *books* to take home	In that library borrow the pupils *bouks* to take home
Paraphrase	In that library the pupils borrow *books* to take home	In that library the pupils borrow *bouks* to take home
Low cloze	De kussens zijn opgevuld met *boeken* waardoor ze hard aanvoelen	De kussens zijn opgevuld met *boekun* waardoor ze hard aanvoelen
Word-by-word translation	The pillows are stuffed with *books* which make them feel hard	The pillows are stuffed with *bouks* which make them feel hard
Paraphrase	The pillows are stuffed with *books* which make them feel hard	The pillows are stuffed with *bouks* which make them feel hard

been reported by Forbes and Connolly (cited in Newmand & Connolly, 2004) and Newman and Connolly (2004). This early ERP, referred to as N270, was claimed to represents a purely orthographically mediated process. Specifically, the amplitude of the N270 would be modulated by mismatches between the orthographic input and the expected orthographic form. They found evidence in some participants for this effect. Because our high-cloze and low-cloze pseudohomophones are orthographically inconsistent with the expected word, we propose that an N270 to pseudohomophones might occur in both the high and low-cloze conditions.

2 Results

2.1 Reaction time pilot study

MANOVAs were performed for the RT and the error data with repeated measures on Cloze (high vs. low) and Lexicality (word vs. pseudohomophone).

For RT a main effect of cloze (F (1, 32) = 51.56, $p < .001$) revealed that mean RT for high-cloze sentences was shorter than for low-cloze sentences (see Table 2). In addition, a cloze by lexicality interaction (F (1, 32) = 14.56, p = .001) reflected that only for the low-cloze sentences mean RT to words was longer than that to pseudohomophones (F (1, 32) = 17.43, $p < .001$). The error analyses revealed a main effect of cloze (F (1, 32) = 6.11, $p < .02$), indicating that participants made less errors on high-cloze sentences (6 %) than on low-cloze sentences (9 %). The cloze by lexicality interaction was marginally significant (F (1, 32) = 3.132, p = .086). The trend reflected that for the low-cloze sentences, more errors were made on the words than on the pseudohomophones (F (1, 32) = 4.11, p = .051). No such difference was found for the high-cloze sentences ($F < 1$).

2.2 Event-related potentials

Grand averages for the high-cloze sentences and for the low-cloze sentences time-locked to the onset of the critical letter strings are presented in Figures 1 and 2, respectively. All conditions elicited the for visual stimuli characteristic early ERP response - that is, an N1 followed by a P2 which at occipital sites was preceded by a

Table 2 Mean reaction time (RT) and error percentages (Error) with standard deviations (SD), for the high-cloze sentences low-cloze sentences and correct words, and pseudohomophones

	High-cloze context				Low-cloze context				Mean RT	Mean error
	RT	S.D.	Error	S.D.	RT	S.D.	Error	S.D.		
Correct Word	804.10	193.21	.06	.06	892.96	196.70	.11	.11	848.53	.08
Pseudo-homophone	811.77	186.36	.06	.05	849.75	176.87	.07	.07	830.76	.06
Mean	807.93		.06		871.35		.09			

The Means are Marginal Means Averaged over either Cloze or Lexicality.

P1. These early components were followed by a broad negative-going wave peaking at about 400 ms, the N400, which was followed by a slow positive shift, the P600, starting at about 500 ms and extending up to 1000 ms. Inspection of Figures 1 and 2 suggests that, (1) words and pseudowords elicited an N400, and (2) pseudohomophones elicited a large P600 (mean amplitudes were more positive for pseudohomophones than for words) in high-cloze sentences but not in low-cloze sentences.

2.2.1. Statistical analyses

The mean percentage of trials that was rejected based on artifacts for the high-cloze condition and the control condition was 26.12 % and 26.03 %, and for the low-cloze and the control condition was 26.91 % and 26.97 %, respectively.

The description of the ERP results will be restricted to main effects and interactions that are relevant for the cognitive-functional interpretation of the condition effects.

2.2.2. N400 window (400-500 ms)

A main effect of cloze reflected that mean N400 amplitude across conditions was larger (i.e., more negative-going) to the critical letter strings in low-cloze sentences than in high-cloze sentences (midline sites: F (1,32)= 48.63, p <.001; lateral sites: F (1,32)= 30.43, p <.001). For the midline sites, a trend for a lexicality effect (F (1,32) = 3.30, p =.079) and a lexicality by site interaction (F (4,29)= 3.12, p <.05) reflected that at posterior sites mean N400 amplitude was more negative for words than for pseudohomophones. For the lateral sites a five-way interaction between cloze, lexicality, ROI, hemisphere and site was found (F(4,29)= 3.07, p <.04). This interaction revealed differences in N400 pattern between words and pseudowords as a function of cloze. Follow up analyses for the high-cloze condition revealed larger N400 amplitudes for pseudohomophones than words at two sites of the left hemisphere (LT and T5: ps <.05). The analyses for the low-cloze condition revealed larger N400 amplitudes for words than pseudohomophones at posterior sites of the right-hemisphere and the left occipital site (P4, P4P, OR, T6, RTP, RT, and OL: all ps <.05), whereas N400 amplitude for a subset of left-hemisphere sites (F7, LAT, and LT: all ps <.02) was larger for pseudowords than for words (see Figure 2).

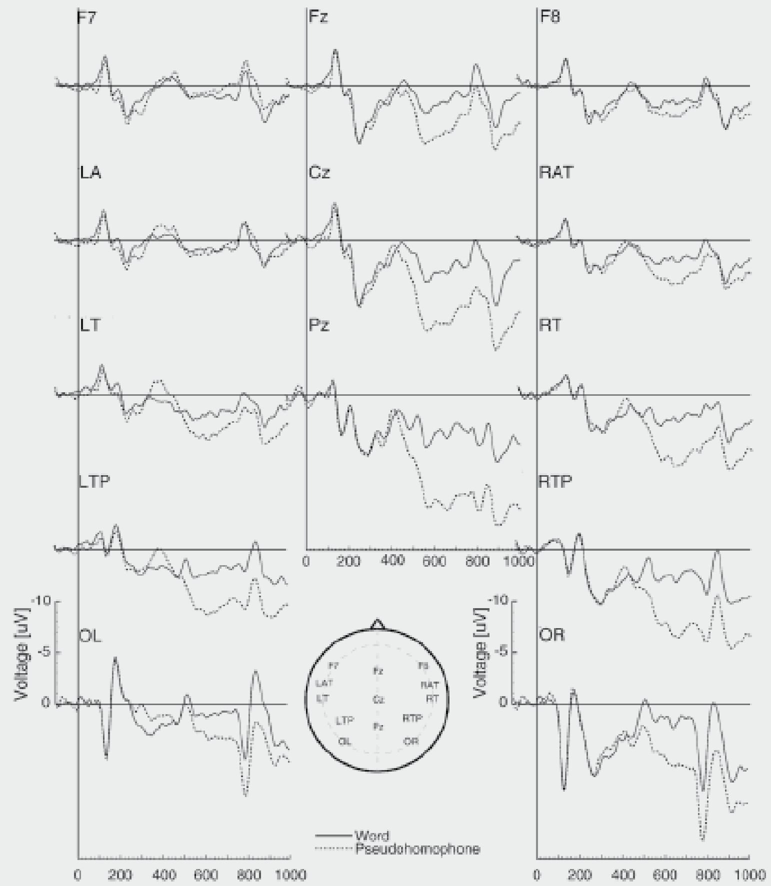
2.2.3 P600 window (650-850 ms)

Effects of cloze for the midline sites (F(1, 32) = 33.18, p < .001) and for the lateral sites (F(1, 32) = 35.63, p < .001), reflected that overall mean amplitudes were more positive for the high-cloze sentences than for the low-cloze sentences. In addition, a cloze by lexicality interaction for the midline sites (F (1,32) = 25.10, p < .001) and for the lateral sites (F (1,32) = 33.20, p < .001) was present. Therefore, separate analyses were performed for the two levels of cloze probability.

2.2.4 High-Cloze Sentences

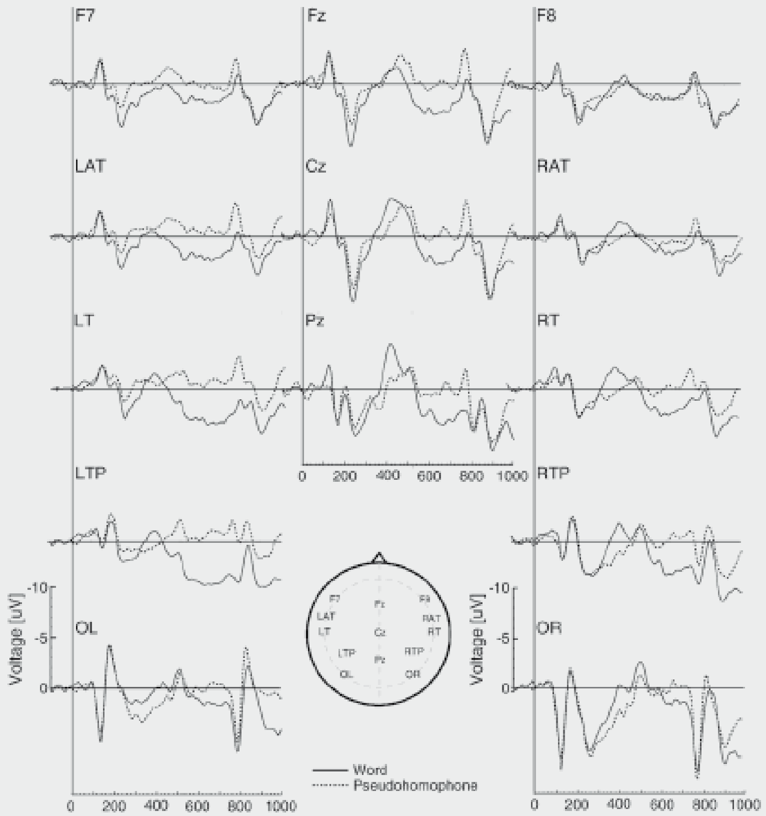
Effects of lexicality confirmed that for the high-cloze sentences mean P600 amplitude was larger for pseudohomophones than for words [midline sites (F (1, 32) = 26.17, p < .001); lateral sites (F (1, 32) = 29.67, p < .001)]. The midline analysis yielded a lexicality by site interaction (F (4, 29) = 14.48, p < .001). Follow up single site analyses revealed larger P600 amplitudes for pseudohomophones than words at Fz, Cz, Pz, and Oz (ps <.05). The lateral analyses revealed interactions between lexicality and hemisphere (F (1,32) = 13.43, p <.01), lexicality and ROI (F (1,32)= 59.67, p <.001), and between lexicality, ROI and site (F (4,29)= 7.81, p <.001).

Figure 1 Grand ERP averages for all midline and a subset of lateral sites, for the high-cloze condition, for words vs. pseudohomophones. Averages are time-locked to the onset of the critical word and superimposed for the two levels of lexicality.



Follow up analyses revealed larger P600 amplitudes to pseudohomophones than words at bilateral temporal and posterior sites (Lt, Rt, Ltp, Rtp, P3, P4, T5, T6, P3p, P4p, Ol, and Or; ps < .01) and at two anterior sites (F4 and Rat; p <.01).

Figure 2 Grand ERP averages for all midline and a subset of lateral sites, for the low-cloze condition, for words vs. pseudohomophones. Averages are time-locked to the onset of the critical word and superimposed for the two levels of lexicality.



2.2.5 Low-Cloze Sentences

Main effects of lexicality [midline sites: (F (1, 32) = 4.93, p < .05); lateral sites: (F (1,32)= 9.28, p <.01)] disclosed that overall mean amplitudes were more negative-going for pseudohomophones than for words. Thus, at first sight a reversed P600

pattern seemed to be elicited by the pseudohomophones in the low-cloze sentences. Closer inspection of the waveforms, however, seems more compatible with a different interpretation, namely, that the unexpected pattern was caused by the unrelated word condition and not by the pseudohomophones. Specifically, the low-cloze word condition at central/posterior sites shows a biphasic pattern –that is, a large N400 followed by a P600 (see e.g., Figure 2: Pz, P4, and RTP). Therefore, in the following the ERP differences will be described as an increase in P600 amplitude to the low-cloze words compared with the pseudohomophones.

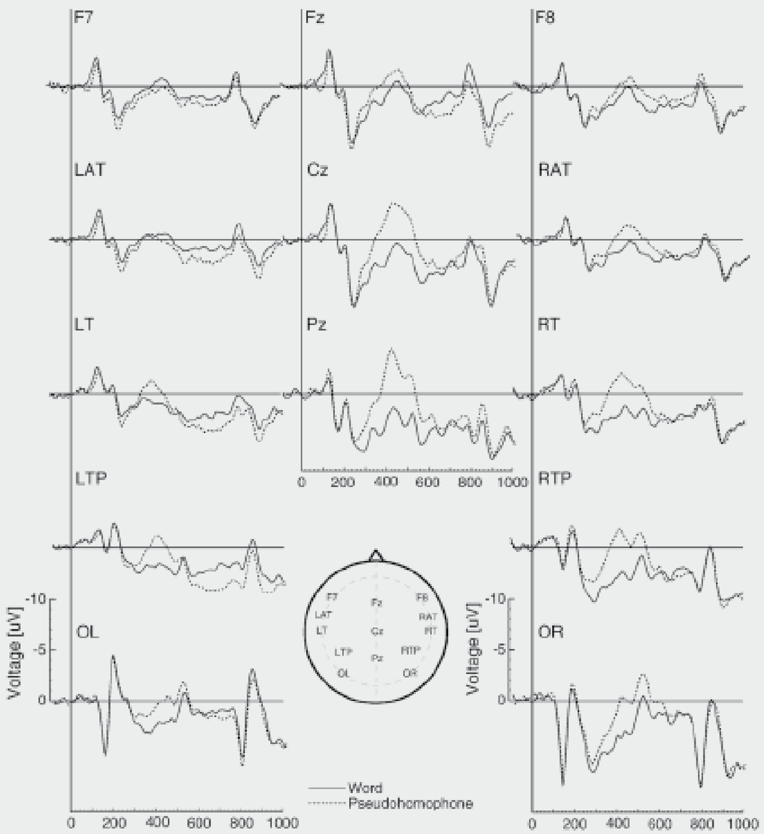
For the midline sites a lexicality by site interaction (F(4,29) = 3.082, p <.05) reflected larger P600 amplitudes for words than pseudohomophones at Fz, Pz, and Oz (all ps < .05). The lateral analyses yielded interactions between lexicality and hemisphere (F (1,32)= 10.54, p <.01), between lexicality and ROI (F(1,32)= 10.91, p <.01), and between. lexicality, ROI and site (F(4,29)= 18.79, p <.001). Separate analyses for the single sites indicated that mean P600 amplitude for words was larger than for pseudohomophones at the following temporal and temporoparietal sites (Lt, Rt, Ltp, Rtp, T5, and T6), bilateral posterior sites (Ol, Or, P3, P4, P3p, and P4p; ps < .01), and two left anterior sites (F3 and LAT; p < .05).

To sum up, the follow up analyses of the cloze by lexicality interactions confirmed that different P600 patterns were obtained for the two levels of cloze probability. As predicted pseudohomophones embedded in high-cloze sentences elicited a large P600 compared to the correctly spelled highly expected word (counterpart). The P600 effect was widely distributed across the scalp and showed a central/posterior maximum. An unexpected result was that the critical words in low-cloze sentences elicited a larger P600 than pseudohomophones.

2.2.6. N400 effect of cloze probability for words and pseudohomophones

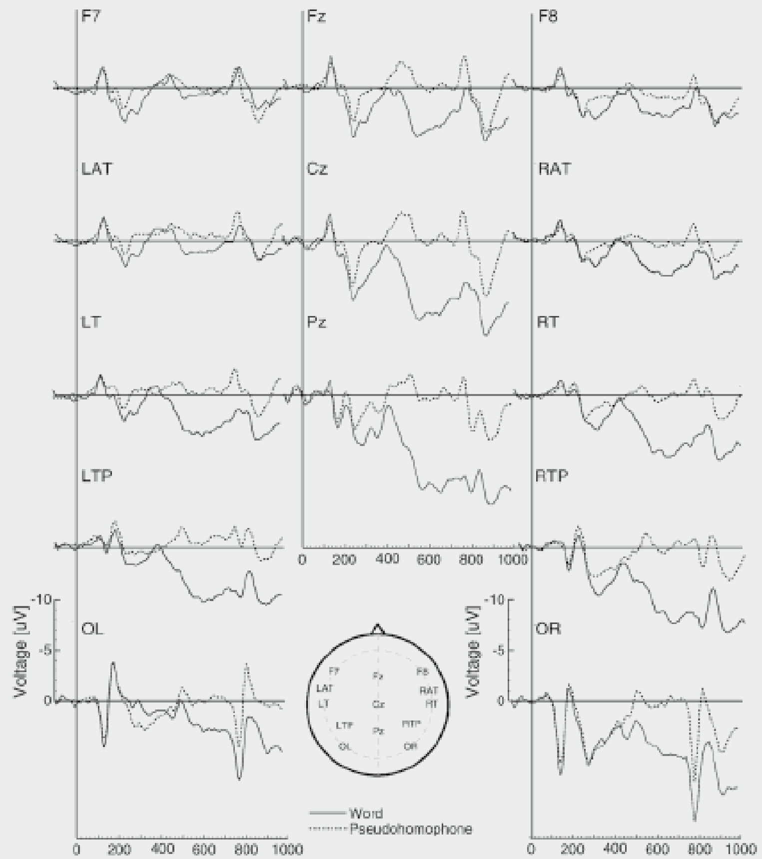
The N400 analyses presented above focused on the comparison of the critical items –that is, words vs. pseudohomophones. The question remained whether a standard N400 effect of cloze probability occurred in the present study (see Figures 3 and 4). Therefore, supplementary analyses were conducted for the words only. These analyses revealed clear effects of cloze for the midline sites (F(1, 32) = 28.23, p < .001) and for the lateral sites (F(1,32) = 26.57, p <.001). A cloze by site interaction for the

Figure 3 Grand ERP averages for all midline and a subset of lateral sites, for the word condition, for high-cloze vs. low-cloze. Averages are time-locked to the onset of the critical word and superimposed for the two levels of cloze.



midline sites (F(4,29) = 9.00, p <.001) indicated that a standard N400 effect was present at central/posterior sites (Cz, Pz, and Oz: ps <.05). The lateral analysis yielded a four-way interaction of cloze by ROI by hemisphere by site, F (4,29) = 3.54, p <.05). The interaction reflected: first, that an N400 effect was present at bilateral

Figure 4 Grand ERP averages for all midline and a subset of lateral sites, for the pseudohomophones, for high-cloze vs. low-cloze. Averages are time-locked to the onset of the critical word and superimposed for the two levels of cloze.



temporal (Lt, Rt, LTP, RTP, T5, and T6: p <.05), and bilateral posterior sites (P3p, P4p, P3, P4, Ol, and OR: p <.05), and, second, that only for the right hemisphere N400 effects extended to anterior/temporal sites (F4a, F4, F8, and Rat: p <.05).

To check whether an N400 effect of cloze was also obtained for pseudohomophones the same set of analyses was carried out for pseudowords only. These analyses also revealed clear effects of cloze for the midline sites ($F(1,32) = 33.94$, $p < .001$) and for the lateral sites ($F(1,32) = 14.67$, $p < .001$). For the midline sites a cloze by site interaction ($F(4,29) = 9.50$, $p < .001$) reflected that – although significant N400 effects occurred at all sites – the N400 effect was most pronounced at Cz and Pz. For the lateral sites an interaction between cloze, ROI, and site was found ($F(4,29) = 11.80$, $p < .001$). Follow up analyses indicated that N400 effects were present for a subset of frontal and temporal sites (F3, F4a, F4, Lt, and Rt: all ps <.05) and for bilateral posterior sites (LTP, RTP, P3p, P4p, P3, P4, and OR: all ps <.05).

In sum, the analyses in which the words and pseudowords were analyzed separately, showed that clear N400 effects of cloze probability were present for words and pseudohomophones. The main difference as a function of lexicality was that the N400 effect to pseudohomophones was more widely distributed than that for words, including frontal midline and left frontal sites.

2.2.7 N270 window

Newman and Connolly (2004) observed an N270 after orthographically incongruent words and pseudohomophones. To check if an N270 was present - like Newman and Connolly - we performed analyses on the most negative peak in the 200 to 350 ms window. These analyses revealed effects of cloze for the midline sites ($F(1, 32) = 7.26$, $p < .05$) and for the lateral sites ($F(1, 32) = 5.67$, $p < .05$). The amplitude of the N270 was larger for low-cloze sentences than for high-cloze sentences. The midline analysis revealed an interaction of cloze by site ($F(4, 29) = 5.01$, $p < .01$). However, separate MANOVAs for the individual sites did not yield reliable effects. More importantly, the lateral analysis yielded interactions of cloze by lexicality by site ($F(4, 29) = 2.95$, $p < .05$) and of cloze by lexicality by ROI by hemisphere ($F(1,32) = 4.44$, $p < .05$). Hence, separate analyses were performed for the two levels of cloze. For the high- cloze sentences no effects or interactions with lexicality were found ($F < 2.5$). For the low-cloze sentences, a lexicality by site interaction was obtained ($F(4,29) = 3.10$, $p < .05$). Follow up analyses indicated that only for the low-cloze sentences N270 amplitude was larger for pseudohomophones than for words at the following left frontal sites (F7, F7a and F3a; all ps < .05).

Discussion

In the Introduction we proposed that a *conflict* triggers a monitoring process in perception, similar to what has been shown in the action domain. This conflict was said to arise between different tendencies: the tendency to reject and the tendency to accept the sentence. Previously, Kolk and colleagues (2003) had claimed that a conflict between two kinds of processing (that is, heuristic and algorithmic processing) triggers a monitoring response and that it is this process that underlies the P600 effect observed to semantically implausible sentences relative to their plausible counterparts.

The prediction of the present study was that the pseudohomophone/high-cloze sentences would lead to a conflict at the word level, between the tendency to accept the pseudohomophone, and the tendency to reject it. The tendency to accept the pseudohomophone was supposed to be very strong not only because it corresponded to a word which was semantically highly expected but also because the phonological form of the word confirmed this expectation and made it maximally strong. On the other hand, the tendency to reject the pseudohomophone would also be very strong, because it is orthographically ill-formed. The resulting conflict was expected to bring the brain into a state of indecision and elicit a monitoring response that should give rise to a P600. In the low-cloze condition, the lexical items from which the pseudohomophones were derived were not expected and thus should not create a mismatch between an expected and an actually presented lexical item. Hence, no monitoring process and no P600 were expected to occur.

The ERP data confirmed the present prediction in that only pseudohomophones embedded in a high-cloze context gave rise to a P600 effect. Because the words from which the pseudohomophones were derived were highly expected, initially the pseudohomophones were easily integrated into the higher order meaning representation of the context. After all, the phonological representation of the pseudohomophone is congruent with the sentential constraints. But when the subject detected the misspelling, which signals a possible processing error, a monitoring response was triggered. This monitoring process gave rise to the P600 effect. Münte et al. (1998) also observed a P600 effect after orthographic anomalies in stories (see Introduction). In the present experiment, we systematically varied the degree of expectation of the critical word by comparing a high-cloze and a low-cloze context. Thereby, in the low-cloze sentences, the words and pseudohomophones were not expected. Consequently, the pseudohomophones were not qualified as possible processing errors and did not confuse the reader; hence, no monitoring process or P600 was triggered. Apparently, a P600 effect is only elicited in cases of a strong conflict, when an unexpected linguistic event is observed while another event is predicted with more or less 100 % certainty.

There was one unexpected finding: namely that a P600 effect followed the N400 effect of cloze probability for the low-cloze words. How can we explain this biphasic pattern? The individual content words that induce the strong expectation for a particular lexical item were the same in the high and low-cloze condition (e.g., '*The pillows are stuffed with feathers*' vs. '*The pillows are stuffed with books*'). Given the high mean cloze value (above 90%), the expectation for a particular noun was very strong. The strong sentence constraint as such could have resulted in a bias to accept the noun, even in the absence of a pseudohomophone that was phonologically identical and orthographically similar to the highly expected word (as was the case for the pseudohomophones in the high-cloze condition). The strong sentence constraint could have resulted in a bias to accept the noun in the low-cloze condition, that is to assume that the highly expected noun was actually presented, to 'fill it in' so to speak. Having actually perceived the low-cloze item, a conflict would arise, leading to reprocessing. This could have resulted in some monitoring activity that gave rise to a larger P600 for the words compared to the pseudohomophones. This explanation also seems in accordance with the RT data. Only for the low-cloze sentences, mean RT to words was longer than that to pseudohomophones [3]. Hence, both the EEG as well as the reaction time data suggest that some monitoring activity occurred to an unexpected event in a strong biasing sentential context.

In line with the literature, a standard N400 effect of cloze probability was found for the high-cloze words versus low-cloze words, when the critical items were in

[3] This opposite to the standard lexicality effect which implies that RTs are shorter to words than to nonwords.

mid-sentence position. Interestingly, an N400 effect of cloze probability was also observed for pseudohomophones when the critical items were in mid-sentence position. This latter finding fits with the proposal by Newman and Connolly (2004) that the phonological representation of the letter string influences the integration of word meaning with sentential context. Because the phonological representation is congruent with the semantic context in the high-cloze sentences but incongruent with the semantic context in the low-cloze sentences, an N400 effect was elicited. The N400 effect for pseudohomophones observed in the present study, speaks in favour of the existence of a phonologically mediated pathway that facilitates semantic integration independently of orthography.

According to Newman and Connolly (2004) an N270 should be elicited whenever a mismatch occurs between a given word and a representation of its orthographic form. In the present study, however, the N270 effect was only observed to pseudohomophones in the low-cloze condition. It is unclear why no N270 was observed to pseudohomophones in the high-cloze sentences. One possibility is that we are dealing with a phenomenon related to what we observed in our sentence processing studies (Kolk et al., 2003). Here, it was found that participants were subject to a temporary semantic illusion, as they did not notice the difference in plausibility between '*the poacher that hunted the fox*' and '*the fox that hunted the poacher*'. In a similar way, our participants may be subject to an orthographic illusion, at least a temporary one, in the high-cloze condition. Just as subjects appear to think that the anomalous sentence is correct in the reversal anomalies, they might think that the misspelled word is orthographically correct in the high-cloze condition. This temporary short-sighted way may be caused by the fact that phonologically and semantically the sentence is intact. It may also be related with the difficulty with proofreading, during which orthographic errors must be detected in otherwise meaningful texts. At some later point, however, the mismatch is detected, leading to a monitoring response, reflected in a P600 effect. In the low-cloze condition, the bias to accept the misspelled word is less strong, and therefore an N270 is elicited. It is clear that future studies are needed to further determine the antecedent conditions for the N270.

The P600 effect has typically been described as an index of syntactic processing (e.g., Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1995). This interpretation has been challenged by studies demonstrating P600 effects after semantic anomalies. Current accounts for a P600 elicited by semantic anomalies propose that individual word meanings ‘cue’, ‘suggest’, or ‘prime’ a plausible role assignment for both plausible and implausible sentences, even in syntactically unambiguous sentences. Furthermore, they assume that the P600 effect is as an immediate consequence of a difference in interpretation: a plausible interpretation on the basis of the individual word meanings and an implausible one by the parser. The accounts differ however in their description of this immediate consequence. One possible consequence of the mismatch between lexical and syntactic analysis was investigated by Van Herten et al. (2005) (see Kim & Osterhout, 2005 for a related idea). They hypothesized that P600 effects to semantically anomalous sentences could arise if the semantic interpretation on the basis of the lexical analysis leads to a strong bias to expect a particular grammatical morphology. The discrepancy between the expected and the observed morphology would then underlie the P600. Van Herten et al (2005) showed that the P600 effect to reversal anomalies was not due to a syntactic mismatch, but was a response to the semantic anomaly (the meaning of the expected verb) as such. A second possibility was suggested by Kuperberg et al. (2003). Since the semantic relationship between the individual words suggests one set of role assignments and the regular parse another, a mismatch occurs. In response to this mismatch, the processing system is said to ‘repair the anomaly by reassigning thematic roles’ (Kuperberg at al., 2003, p. 128). This repair process is of a syntactic nature because it involves a process of restructuring. The notion of reprocessing embodied in this hypothesis certainly seems part of the picture. Syntactic reprocessing is also assumed to occur in garden path sentences, where it serves to uncover the alternative parse and interpretation (e.g., Friederici, 1995). Our view on the P600 is that it reflects reprocessing indeed but that its function is more general than syntactic repair. Its function is to monitor for processing errors. As a consequence, it could involve reprocessing at a number of linguistic levels, just as speech repair may involve phonological, syntactic, lexical and conceptual levels, leading to phonological, syntactic, lexical and appropriateness repairs, respectively. In support of this approach, the present study has demonstrated that in language perception, in addition to a monitoring process at the conceptual/semantic level, there is also a process of monitoring for errors on the orthographic level.

As shown by previous studies, the P600 effect to syntactic and certain semantic violations shows a central/posterior scalp distribution (Coulson, King, & Kutas, 1998; Kolk et al., 2003). The P600 to pseudohomophones had a slightly different scalp distribution. We observed that the present P600 was centroparietally distributed across the scalp extending to two right anterior sites. One might argue that this could indicate that the P600 to pseudohomophones is qualitatively different to the earlier reported P600. However, it must be realized that there is evidence for quite some variation in scalp distribution of the P600 after syntactic and semantic anomalies. A more frontal/broad distribution of the P600 effect has for example also been reported for locally ambiguous sentences (Friederici, Hahne, & Mecklinger, 1996; Hagoort, Brown, & Osterhout, 1999; Osterhout & Holcomb, 1995; Van Berkum, Brown, & Hagoort, 1999) while Kaan and Swaab (2003) observed a more posterior distribution. As Figure 1 shows, the present P600 had an early onset. With respect to P600 latencies, Friederici (1995) has proposed that differences in latency may reflect the complexity of processing necessary for the revision of the initially preferred reading. Longer latencies seem to be correlated with more complex restructuring. An early positivity, referred to as P345 has been observed after disambiguating auxiliaries (Friederici, 1995). Intermediate positivities have been shown to occur after syntactic and certain semantic anomalies (time range: 650-850 ms). Finally, even later positivities (time range: 700-1300 ms) have been reported in a recent discourse study (Nieuwland & Van Berkum, 2005). Hence, differences in latency may indeed be systematically related to the complexity of the language input that has to be checked for possible processing errors.

From this, we conclude that the reported variability in scalp distribution and latency of the P600 does not imply qualitatively different processes. According to the monitoring hypothesis, this variability is understandable. This could be understood as a function of type of material that needs to be processed and its complexity.

It thus seems that, when we are looking at P600 effects of syntactic violation or syntactic ambiguity, we are not observing manifestations of syntactic processing as such, but of higher order – executive – processes of conflict resolution. This conclusion bears a striking resemblance to the one reached by Thompson-Schill and her colleagues, on the basis of fMRI research (for a review see Novick, Trueswell, & Thompson-Schill, 2005). One essential finding is that garden path sentences, which as we saw above reliably elicit P600 effects, also lead to activity in the left inferior frontal gyrus (LIFG). This is not – in view of the authors - because this area is specialized in syntactic processing, but because this is an area for conflict resolution. One important argument is that this area becomes routinely activated in incongruent trials during the Stroop task. Another argument is that damage to just this area leads to only a very minor and transient language impairment, and not to something like agrammatism, as one might expect. On the other hand, patients with LIFG damage have severe difficulty in completing sentences when there are many competing possibilities. It has to be noted that the authors assume that both the LIFG and the Anterior Cingulate are involved in monitoring and conflict resolution, but that they do so under slightly different circumstances. We may conclude that both ERP and fMRI studies point to the presence of higher order processes of cognitive control and monitoring during language comprehension.

4 Experimental procedures

4.1 Participants

Thirty- three students (mean age = 22 years; age range = 18 to 33) participated in the experiment. All were native speakers of Dutch, had no reading disabilities, were right-handed, and had normal or corrected-to-normal vision. Hand dominance was assessed with an abridged Dutch version of the Edinburgh Inventory (Oldfield, 1971). Seven participants reported the presence of left-handedness in their immediate family.

4.2 Materials

We first constructed 127 simple declarative sentence fragments and used these in a cloze test with 25 subjects to obtain highly expected ('high-cloze') critical words. Of these 127 sentences, 116 sentences were completed with the same word by 91% of the participants. These were used as the high-cloze context sentence fragments in this study.

We then created 116 low-cloze context sentences by exchanging the critical word from a high-cloze context fragment with the critical word from another high-cloze context fragment. For example, we exchanged the critical word from 'In that library the pupils borrow *books* to take home.' with the critical word from 'The pillows are stuffed with *feathers* which makes them feel soft.' resulting in the following low-cloze fragment 'The pillows are stuffed with books which makes them feel hard.' The critical word was always in mid-sentence position.

A further experimental manipulation was Lexicality (correct word vs. from the correct word derived pseudohomophone). Every critical word occurred in a correct version and a pseudohomophonic version. The pseudohomophone was created by changing the vowel of the second syllable, keeping phonology the same. Every noun contained two syllables and that the changed vowels were always in the second part of the word.

The two experimental manipulations Context and Lexicality were crossed. As a result there were four conditions and thus, four experimental sentence types: high-cloze correct word sentences, high-cloze pseudohomophone sentences, low-cloze correct word sentences and low-cloze pseudohomophone sentences; yielding a total set of 464 sentences. The four versions of each sentence were counterbalanced across lists. Each list contained each sentence context (in a high-cloze or a low-cloze version) and each critical word (in a correct word or a pseudohomophone version) only once. So, each list contained 29 high-cloze correct word sentences, 29 high-cloze pseudohomophone sentences, 29 low-cloze correct word sentences, and 29 low-cloze pseudohomophone sentences. To each list 60 filler sentences were added: 30 correct sentences, 10 sentences with a pseudohomophone at the beginning of the sentence, 10 sentences with a pseudohomophone in the middle of the sentence, and 10 sentences with a pseudohomophone at the end of the sentence. To avoid sentence wrap-up processes to affect our measurement, every critical word was presented at a mid-sentence position.

4.3 Procedure

For the EEG study, participants were seated in an experimental room. Sentences were presented in serial visual presentation mode at the center of a PC monitor.

Word duration was 345 ms and the stimulus-onset asynchrony (SOA) was 645 ms. Sentence final words were followed by a full stop. The inter-trial interval was 2 seconds. Words were presented in black capitals on a white background in a 9 cm by 2 cm window at a viewing distance of approximately 1 m. Each sentence was preceded by a fixation cross (duration 510 ms) followed by a 500 ms blank screen. The experimental list was split up into five blocks; there was a brief pause between blocks and each block was preceded by two filler items. Participants were instructed to attentively read the sentences. Because eye movements distort the EEG recording, participants were trained to make eye movements, e.g. blinks, only in the period between the end of the last sentence and the beginning of the next sentence.

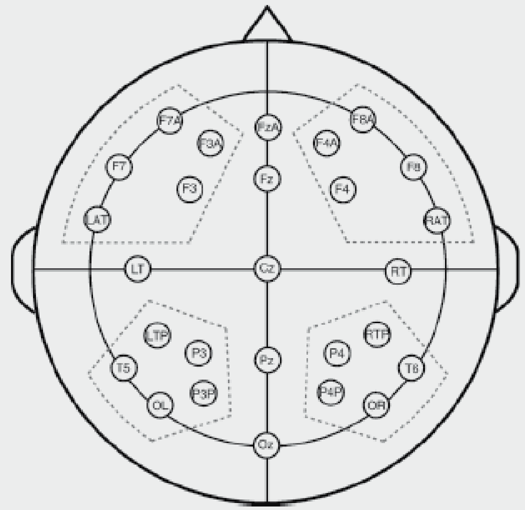
Prior to this EEG study, a reaction time (RT) pilot study was conducted as a pretest of the material, to test if participants were successful in detecting the pseudo-homophones. A separate group of 33 participants was tested that fulfilled the same criteria as those in the ERP study. The procedure differed in two aspects from the ERP study: First, the critical letter string was presented in sentence-final position (e.g. paraphrase: “In that library the pupils borrow *books*.”) and, second, participants performed a lexical decision task. They had to indicate as fast as possible by pressing a button with the right or left index finger if the critical letter string was a real word (right-hand response) or not (left-hand response). A response device with three push-buttons was placed on a small table in front of the participant.

4.4 Electrophysiological recording

The electroencephalogram (EEG) was recorded with 27 tin electrodes mounted in an elastic electrode cap (Electrocap International; see Figure 5 for the montage).

The electrode positions included standard International 10-20 system locations over the left and right hemispheres at the frontal (F3, F4, F7, and F8), midline (Fz, Cz, Pz, and Oz), parietal (P3 and P4) and temporal (T5 and T6) sites. Eight extra electrodes were placed at the frontal (F3A, F4A, F7A, and F8A), midline (Fza and Oz) and parietal (P3P and P4P) sites. In addition, eight electrodes were placed at non-standard electrode positions previously found to be sensitive to language manipulations (e.g. Holcomb and Neville, 1990): left and right anteriortemporal sites

Figure 5 Electrode configuration used in the present experiment.



(LAT and RAT: 50% of the distance between T3/4 and F7/8), left and right temporal sites (LT and RT: 33% of the interaural distance lateral to Cz), left and right temporoparietal (LTP and RTP: Wernicke’s area and its right hemisphere homologue: 30% of the interaural distance lateral to a point 13% of the nasion-inion distance posterior to Cz), and left and right occipital sites (OL and OR: 50% of the distance between T5/6 and O1/2). The left mastoid served as reference. Electrode impedance was less than 3 KΩ. The electro-oculogram (EOG) was recorded bipolarly; vertical EOG was recorded by placing an electrode above and below the right eye and the horizontal EOG was recorded via a right to left canthal montage. The signals were amplified (time constant = 8 s, bandpass = 0.02 – 30 Hz), and digitized online at 200 Hz. Presentation of stimuli and recording of performance data was accomplished by a Macintosh computer.

4.5 Data analyses

Before analysing EEG and EOG, records were examined for artifacts and for excessive EOG amplitude (>100 μV) from 100 ms before the onset of the critical letter string ending the relative clause to 1000 ms following its onset. Averages were aligned to a 100-ms baseline period preceding the critical letter string. Based on visual analysis and previous studies (Kolk et al., 2003; Van Herten et al., 2005), mean amplitudes in the 400-500 ms and 650-850 ms window were taken as N400 window and P600 window, respectively.

For the two time windows, repeated measures Analyses of Variance (ANOVAs) were conducted separately for the midline sites and for the lateral sites with cloze (high vs. low) and lexicality (word vs. pseudohomophone) as factors. The midline analyses included the additional factor site (Fza, Fz, Cz, Pz, Oz). To explore the scalp distribution of the ERP effects for the lateral analyses we used a region of interest (ROI: anterior vs. posterior) by hemisphere by lateral site (F7a/F3a/F7/F3/LAT vs. LTP/P3/P3p/T5/OL vs. F8a/F4a/F8/F4/RAT vs. RTP/P4/P4p/T6/OR) design. The multivariate approach to repeated measurements was used to avoid problems concerning sphericity (e.g. Vasey and Tayer, 1987). Interactions with the factor site were followed up by single site analyses.

Acknowledgments

Portions of this research were presented at the 12th Annual Meeting of Cognitive Neuroscience Society in New York, 2005. We thank Gina Kuperberg and an anonymous reviewer for their very helpful comments on an earlier version of this article. We are grateful to the ERG group of the NICI for technical assistance and to Maarten vd Meulen for his support.

References

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624-652.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed processing approaches. *Psychological Review*, 100, 589-608.

Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilisitc word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117-1121.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology,* 47, 164-203.

Ferreira, F., Ferraro, V., & Bailey, K.G.D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.

Friederici, A.D. (1995). The time-course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, 50, 259-281.

Friederici, A.D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1219-1248.

Gunter, C., Friederici, A.D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience,* 12, 556-568.

Hagoort, P., Brown, C.M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes,* 8, 439-483.

Hagoort, P., Brown, C.M., & Osterhout, L. (1999) The neurocognition of syntactic processing. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273-317). Oxford: University Press.

Hartsuiker, R.J., & Kolk, H.H.J. (2001). A computational test of the perceptual loop theory. *Cognitive Psychology*, 24, 113-157.

Hoeks, J.C.J., Stowe, L.A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59-73.

Holcomb, P.J., & Neville, H.J. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes,* 5, 281-312.

Kaan, E., & Swaab, T.Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience,*15, 98-110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language,* 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language,* 85, 1-36.

Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P.J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes,* 21, 489-530.

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research,* 17, 117-129.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition,* 14, 41-104.

Münte, T.F., Heinze H.-J., Matzke M., Wieringa, B.M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia,* 36, 217-226.

Newman, R.L., & Connolly, J.F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Cognitive Brain Research,* 21, 94-105.

Nieuwland, M.S., & Van Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research,* 24, 691-701.

Novick, J.M., Trueswell, J.C., & Thompson-Schill, S.L. (2005). Cognitive control and parsing: Reeaxamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience,* 5, 263-281.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia,* 9, 97-113.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language,* 31, 785-806.

Osterhout, L., & Holcomb, P.J. (1995). Event-related brain potentials and language comprehension. In M.D. Rugg & M.G.H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp171-215). New York: Oxford University Press.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition,* 77, 97-131.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language,* 41, 147-182.

Van Herten, M., Kolk, H.H.J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research,* 22, 241-255.

Van Herten, M., Chwilla, D.J., & Kolk, H.H.J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitve Neuroscience*, 18, 1181-1197.

Vasey, M. W., & Tayer, J. F. (1987).The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution, *Psychophysiology*, 24, 479-486.

Yeung, N., Botvinick, M.M., & Cohen, J.D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111, 931-959.

Chapter 4

Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task[1]

[1] This chapter has been published as Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk, H.H.J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967-982.

Abstract

P600 effects have been observed after syntactic ambiguous sentences, after several types of syntactic and semantic anomalies and after orthographic anomalies. On the basis of these findings, several investigators propose the P600 effect to reflect syntactic repair or syntactic restructuring. According to our Monitoring Theory the P600 effect reflects more general sentence reanalysis, to check whether the input sentence has been perceived appropriately. When the brain encounters a highly unexpected linguistic event, a conflict arises between the expected representation and the representation derived from the input. This conflict is proposed to trigger a process of reanalysis. In the present study, expectancy was manipulated by varying the truth-value of a sentence in relation to a picture. ERPs were recorded from 27 electrodes while we presented participants (N= 30) pictures of spatial arrays followed by a sentence giving a correct or incorrect description of the picture. The mismatches were predicted to create a conflict between the conceptual representation on the basis of the picture and the actual sentence and should therefore lead to a P600. A P600 effect was indeed observed after both intra-dimensional □ △ - '*the triangle stands in front of the square*.', and extra-dimensional □△ - '*the triangle stands below the square*.' mismatches. The present results support our Monitoring Theory; that is, the function of reprocessing reflected by the P600 effect is not purely syntactic repair or restructuring but is more general in nature, to check for possible processing errors.

Introduction

From the perspective of a reader trying to understand a sentence, syntactic ambiguity is a challenge, often leading to errors of comprehension and to syntactic reanalysis. For instance, readers of so called 'garden path' sentences which contain temporary syntactic ambiguities like '*the woman persuaded to answer the door*', chose first for an active interpretation of the verb *persuade* (and thus assume that the sentence is about a woman persuading someone), expect an object NP after the verb, but read *to* instead. This forces them to reanalyze the sentence, in order to arrive at an interpretation in which the verb is taken as a passive participle (and thus realize that the sentence is about a woman being persuaded). So, there is a syntactic bias in these garden path sentences: that is, a strong preference for a particular structural analysis, which turns out to be the wrong one later on in the sentence. Numerous studies in which eye movements were recorded during the presentation of sentences containing such structural ambiguities, showed that readers experience difficulty reading the disambiguating material; that is, these studies generally found that eye fixations are longer and/or eye regressions become more frequent when readers encounter the disambiguating material which is inconsistent with the preferred structure (Frazier & Rayner, 1982; Rayner et al., 1983; see Frazier, 1987, for a review).

In the ERP domain, a P600 effect was reliably observed after garden path sentences (e.g., Friederici, Steinhauer, & Frisch, 1999; Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994). This P600 Event Related Potential (ERP) is a late positive-voltage effect starting at about 600 ms after the onset of a target word. A process of syntactic reanalysis or repair is assumed to be responsible for the occurrence of this P600 effect. Amplitude and latency of the P600 are thought to vary as a function of the difficulty of recovery of the garden path (Friederici & Mecklinger, 1996; Osterhout et al., 1994). Difficulty of recovery is proposed to be correlated with the degree of complexity of the required syntactic reanalysis. More complex reanalysis processes are accompanied with longer latencies. The proposed function of the P600 after syntactically correct but temporarily ambiguous sentences is to revise the syntactic structure and uncover the alternative parse; that is, to replace one sentence parse by another (Friederici, 1995).

Although in these studies the P600 effect has been interpreted as a brain response to ambiguity, P600 effects have also been observed in response to syntactically unambiguous sentences. A P600 effect is for example observed after different types of semantic anomalies; this has been shown in different languages (Kolk, Chwilla, van Herten, & Oor, 2003; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003, Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Van Herten, Kolk, & Chwilla, 2005; Van Herten, Chwilla, & Kolk, 2006) [2]. For instance, Kim and Osterhout (2005) observed a P600 effect after anomalous verbs in simple active sentences, for example after '*devouring*' in the sentence '*The hearty meal was devouring* …'. Hoeks et al. (2004) also observed a P600 effect after semantic verb-argument violations without syntactic ambiguities, such as after '*thrown*' in the sentence '*the javelin has thrown the athletes*'. The question arises why the P600 effects could be reliably produced by the above described semantic violations since these sentences were syntactic unambiguous. For the semantic violations, the above proposed function of syntactic reanalysis is less obvious, because the unambiguous syntactic structure allows only a single parse.

What then could be the source of the P600 effect after semantic violations? One possibility is that these semantic anomalous sentences involved some kind of syntactic anomaly. It is well-known that syntactic anomalies give rise to P600 effects. An increase in P600 amplitude has for example been observed after subject-verb agreement violations (Hagoort, Brown, & Groothusen, 1993; Osterhout & Mobley, 1995; Vos, Gunter, Kolk, & Mulder, 2001) or after phrase structure violations (Neville, Nicol, Barss, Forster, & Garett, 1991; Hahne & Friederici, 1999; Friederici,

[2] These findings came as a surprise since semantic anomalies (e.g., 'the cat will bake the food') were known to elicit a negative wave that peaks at about 400 ms after an anomalous word; that is, the N400 effect (e.g., Kutas & Hillyard, 1980, 1984; Osterhout & Nicol, 1999). The N400 effect is described as an index of semantic integration, in that it reflects the ease with which a word is integrated into its context, be this a single word (Chwilla, Brown, & Hagoort, 1995; Chwilla, Hagoort, & Brown, 1998; Holcomb, 1993), a sentential context (Friederici, 1995; Van Petten & Kutas, 1990) or a discourse context (St. George, Mannes, & Hoffman, 1994; Nieuwland & Van Berkum, 2005).

Pfeifer, & Hahne, 1993). Now, the semantic anomalies do not contain an explicit syntactic violation. However, Van Herten et al. (2005) considered the possibility that participants would expect a particular inflection in these sentences which would then fail to match with the observed inflection. In particular, if participants would combine the lexical items of the sentence '*The cat that fled from the mice* …' in the most plausible way, they could assume that the mice were fleeing from the cat. This interpretation could lead them to expect a particular inflection of the verb (here: a plural inflection). The violation of this expectation could have elicited the P600 effect. Such a syntactic mismatch can only occur in sentences in which the number of theme and agent are different. However, Van Herten et al. (2005) found a P600 effect in both different number sentences but also in same number sentences. Consequently, the P600 effect cannot be due to a syntactic mismatch. Another possible source of the P600 effect after semantic violations could be syntactic complexity. The P600 effect was shown to differ as a function of syntactic complexity; specifically, an increase in P600 amplitude has been observed after a high degree of sentence complexity (Kaan, Harris, Gibson, & Holcomb, 2000). However, the semantic violations were no more syntactically complex than their control sentences.

In the above, we indicated that a process of syntactic reanalysis as has been proposed for garden path sentences is not likely to occur for our semantic anomalies, as these sentences are not ambiguous. Nevertheless, a process of syntactic reanalysis has been suggested also for these sentences (see for example Kuperberg et al., 2003; Kim & Osterhout, 2005). Here, reanalysis has a somewhat different form, however; rather than a complete reparse of the sentence, it is proposed that the thematic roles are reassigned. In many types of semantic anomalies evoking a P600 rather than an N400, there was a strong semantic-thematic 'attraction' (Kim and Osterhout, 2005) or a potential semantic-thematic 'fit' (Kuperberg et al., 2003) between the critical verb and its preceding argument. Specifically, the semantic anomalies were verb-argument semantic violations in which the arguments could have occupied alternative thematic roles which are more plausible. Many of these anomalous sentences were, therefore, repairable by reassigning the thematic roles of the critical verb's arguments to the more plausible thematic roles. For example, sentences like '*Every morning at breakfast the eggs would eat*', would make sense if the reader would reassign the agent role of '*eggs*' to a theme role, such as in '*Every morning at breakfast the eggs would be eaten*'. So, the P600 effects observed after semantic anomalies with a strong semantic fit or attraction could represent syntactic reanalysis; involving an online attempt to structurally repair a sentence by reassigning thematic roles [3].

However, we question this proposed function of reanalysis after such violations. To account for the P600 effect after *'at breakfast the eggs would eat'*, one could assume as described above, that readers replace an implausible role assignment with a plausible one. But unlike in garden path sentences, in this syntactically unambiguous sentence there is nothing to replace because the syntactic structure allows only one role assignment. Becoming involved in restructuring would lead participants to assume the most plausible but incorrect role assignment; thematic role reassignment would thus be dysfunctional. And in fact, there is no evidence that participants actually do replace an implausible role assignment with a plausible one because they almost always classify the sentence correctly in a judgment task (e.g., Kuperberg et al., 2003; Kolk et al., 2003).

Perhaps, a process of syntactic reanalysis does occur after semantic anomalies but for a different reason. Perhaps its function is not to reassign roles but to reanalyze the sentence completely, in order to find out whether one has read the sentence correctly. This is the essence of the Monitoring Theory we have proposed (see also, Kolk et al., 2003; Van Herten et al., 2006; Vissers, Chwilla, & Kolk, 2006, 2007). We claim that the *functional* interpretation of the P600 should be extended from a purely syntactic reanalysis account to a more general sentence reanalysis account. We explain this more broad function as follows. While reading a text or listening to a conversation, we can make errors in perception (e.g., misreading a word). Since integration of the results of such misperceptions would sincerely endanger discourse coherence, it seems likely that these perceptual errors are edited out. The function of reanalyzing or monitoring in language perception would then be to find out whether the cause of a particular inconsistency could be a processing error,

[3] To be sure, the view of Kuperberg and her colleagues have evolved and they now put more emphasis on notions like integration costs and on factors determining the likelihood that an anomaly will be detected (Kuperberg, 2007).

due to mishearing or misreading. By way of illustration, it is like questioning oneself 'Did I hear or read that correctly?' After all, an inconsistency can have two sources. It can be real, in the sense that an unexpected event did indeed occur. On the other hand, it can also stem from a processing error. Monitoring will not set in after encountering any anomaly, but only when there is a high degree of uncertainty about the source of the anomaly: that is, the less an event is expected, the more chance there is that the event is due to erroneous processing.

To summarize, according to the Monitoring Theory the P600 does reflect reanalyzing as has been proposed by several other investigators. This reanalyzing can also be of a syntactic nature, though not necessarily so. The main difference between the Monitoring Theory and other reanalysis accounts concerns the *function* of the reprocessing. We claim that the reanalyzing does not serve to select a different parse of an ambiguous sentence, as proposed by Friederici (1995). Nor does reanalysis serve to obtain a new parse or set of thematic roles after a violation of grammatical-semantic constraints (e.g., Kuperberg et al. 2003); after all, in the syntactic unambiguous, semantic anomalies, a different parse would be impossible and a different role reassignment would be dysfunctional. Instead, we claim that reprocessing serves to check whether the initial sentence processing has been correct. We state that the P600 reflects a process of reanalysis in language perception; aimed at prevention of integration of false information into the discourse representation.

Now, the question becomes: how does the monitoring system detect a possible processing error? Our proposal is that it is a conflict between different analyses of the same linguistic string that triggers the reanalysis. That is, in a garden path sentence, the active interpretation which is strongly preferred on the basis of the syntactic bias conflicts with the less preferred correct interpretation in which the verb is taken as a passive particle. We thus propose that in these sentences, the activation of two incompatible interpretations leads to a conflict; this conflict triggers a process of reanalysis and hence a P600, to check for the possibility of a processing error. Likewise, the Monitoring Theory proposes that the P600 after semantic anomalies reflects reanalysis triggered by a conflict between two kinds of processing; that is, between semantic and syntactic processing. In particular, when reading the sentence '*the fox that hunted the poachers stalked through the woods*', plausibility heuristics and syntactic algorithms run in parallel and produce different thematic interpretations. Whereas the plausibility heuristic leads to the most plausible interpretation of the set of content words that occur in the sentence (a lexical strategy), that is that poachers are hunting foxes, the parsing routines lead to the interpretation that the foxes are hunting the poachers. That readers actually pursue a word-based analysis is supported by the fact that the usual ERP index of semantic anomaly, the N400 effect, was absent after semantic reversal anomalies. The observed P600 effect is assumed to reflect an immediate consequence of the fact that the parse and the plausibility heuristic suggest different interpretations. Most importantly, the proposed function of reanalysis is the same for both sentence types; that is, resolving the conflict by monitoring the input for possible processing errors.

The Monitoring Theory puts forward that the P600 reflects reprocessing with the more general function of checking whether the initial sentence processing has been correct. As a consequence, it could involve reprocessing at a number of linguistic levels. In support of this, we have demonstrated that in addition to a monitoring process at the sentence level, there is also a process of monitoring for errors on the word level (Vissers et al., 2006). That is, a P600 effect was observed after pseudohomophones in a high cloze context (e.g., '*In that library the pupils borrow bouks to take home*'), whereas the same pseudohomophones in a low cloze context did not evoke a P600 effect. Important for the present argument, the P600 effect after these orthographic violations cannot be due to rearranged thematic roles (e.g., Kuperberg et al., 2003); since these anomalies are purely orthographic, thematic reassignment would not be helpful. We explained the observed P600 effect as follows. In the high cloze condition, the tendency to accept the pseudohomophone was very strong because it corresponded to a word which was semantically highly expected and also because the phonological form of the word confirmed this expectation and made it maximally strong. On the other hand, the tendency to reject the pseudohomophone was also very strong, because it was orthographically ill-formed. So, the pseudohomophone/high cloze sentences lead to a conflict at the word level, between the tendency to accept the pseudohomophone, and the tendency to reject it. This conflict between response tendencies gave rise to a process of reanalysis which led to the P600 effect. Hence, in support of the Monitoring Theory, we have demonstrated that in language perception a process of monitoring for errors can take place at a number of linguistic levels.

The purpose of the present experiment was to further test the Monitoring Theory by presenting sentences, the meaning of which could either match or mismatch a preceding picture. Such a sentence-picture matching task was also used in investigating on-line thematic role assignment by Wassenaar and Hagoort (2007). They presented subjects with a picture that was followed by a syntactically correct sentence. The thematic roles of the sentence either matched or mismatched the thematic roles displayed in the picture. For example, in the mismatch condition, after presentation of a picture in which a woman pushes a man in a wheelchair, the sentence '*the tall man on this picture pushes the young woman*' is presented. A P600 effect was observed after the sentences that did not fit the depicted thematic roles. Wassenaar and Hagoort propose that in these sentences the role assignment based on the picture interferes with the role assignment based on the sentence; since the picture indicates one role assignment, it is difficult to assign the reverse thematic roles in the sentence. According to this interference hypothesis, the size of the P600 effect varies as a function of how effortful the assignment process is. The Monitoring Theory accounts for the P600 effect in a different way, by proposing that it reflects a monitoring response triggered by the conflict between the predicted thematic roles on the basis of the picture representation and the thematic roles of the sentence.

In the present experiment, we present syntactically correct and unambiguous sentences which correctly or incorrectly describe a preceding picture with geometrical figures. In particular, a picture of a spatial array is shown, for instance of a square in front of a triangle. Then a sentence is presented which can either match or mismatch with the picture, depending on the preposition being used. Two different types of mismatch are used; an intra-dimensional mismatch (e.g, *behind* in the picture, *before* in the sentence) and an extra-dimensional mismatch (e.g., *before* in the picture, *above* in the sentence).

□△	De driehoek staat achter het vierkant *The triangle stands behind the square*	(match)
□△	De driehoek staat voor het vierkant *The triangle stands in front of the square*	(intra-dimensional mismatch)
□△	De driehoek staat boven het vierkant *The triangle stands above the square*	(extra-dimensional mismatch)

By employing locative relationships, we avoid the involvement of thematic role assignment and thereby a possible effect of role assignment interference, as proposed by Wassenaar and Hagoort (2007). This is because thematic roles refer to actions, in particular to the way different entities are involved in such actions. So, in the case of locative relationships we are not dealing with arguments.

However, one may still argue that location embodies a specific type of thematic role. On this account one could maintain that for an intra-dimensional mismatch like '*the triangle stands in front of the square*.', where the picture presents the reverse situation, '*square*' and '*triangle*' have switched roles, and that this switching could be repaired by role assignment. To definitively rule out a potential role assignment explanation, the extra-dimensional mismatches were added. These mismatches are purely semantic and do not involve role switches in any possible way. So, re-assigning roles would not be helpful in this condition. Another distinguishing feature of the nouns used in the present study is that they are all inanimate; so that animacy constraints cannot bias towards a role assignment which is in conflict with a role assignment dictated by the syntax. Animacy can thus not propose strong Agent roles for certain nouns; that is, both nouns can equally likely play a role here (see Packynski, Kreher, Ditman, Holcomb, & Kuperberg, 2006).

We assume that, when faced with a picture, one immediately forms some kind of mental representation of the depicted event. There is strong eye movement evidence that real-time spoken language comprehension takes relevant visual context immediately into account (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). So, pictures will immediately be integrated during the process of language comprehension. This will enable subjects to detect a picture-sentence mismatch when the mental representation on the basis of the picture is in conflict with the sentence representation. Furthermore, we can assume that a sentence is typically expected to be true. So, if a picture is presented, there will be a strong expectation to subsequently read a sentence, which gives a true description of the picture. A sentence giving a false description, on the other hand, is highly unexpected.

Our prediction is that both intra-dimensional and extra-dimensional mismatches will create a conflict between the representation emanating from the picture and the representation derived from the sentence and will therefore elicit a P600 effect.

2 Methods

2.1. Participants

Thirty students (mean age = 21; age range = 18-29) participated in the present experiment. All participants were native speakers of Dutch, had no language disability, had no neurological or psychological impairment, had normal or corrected-to-normal vision, and were right-handed. Handedness was assessed with an abridged Dutch version of The Edinburgh Inventory (Oldfield, 1971).

2.2 Materials

The stimuli for the present experiment consisted of 288 picture-sentence pairs. Half of these picture-sentence pairs were experimental trials and the other half were used as filler trials. The pictures consisted of black on white line drawings depicting two abstract objects (a square, a circle, a triangle, or a star). For half of these pictures the two abstract objects were depicted in the horizontal dimension, and for the other half they were depicted in the vertical dimension. The four abstract objects used resulted in a total of six object pairs. By counterbalancing the position of the objects in the picture, for both dimensions (horizontal and vertical) twelve different pictures were created. The experimental sentences were of the following form: NP_1 VP NP_2. The verb that was used was *staan* (to stand) and the prepositions used were *voor* (in front of), *achter* (behind) (both used in the horizontal dimension), *boven* (above), *onder* (below) (both used in the vertical dimension). NP_1 and NP_2 were counterbalanced and consisted of a determiner and an abstract object (e.g., *de cirkel* (the circle)).

The experiment consisted of three sentence type conditions: (1) match, (2) intra-dimensional mismatch and (3) extra-dimensional mismatch. A sentence matched a picture if it correctly described what was depicted in the picture: i.e., the depicted picture (e.g., □△) matched the presented sentence (e.g., *De driehoek staat achter het vierkant- The triangle stands behind the square*) A picture-sentence mismatch occurred if the sentence description did not match what was depicted in the picture. Mismatches were formed by changing the preposition. To create the intra-dimensional mismatches, the opposite preposition of the same dimension (horizontal or vertical) as compared to the matching condition was used: i.e., the depicted picture (e.g., □△) did not match the presented sentence (e.g. *De driehoek staat voor het vierkant-The triangle stands in front of the square*.) To create the extra-dimensional mismatches a preposition from the other dimension as compared to the matching condition was used: i.e., the depicted picture (e.g., □ △) did not match the presented sentence (e.g., *De driehoek staat onder het vierkant.- The triangle stands below the square*).

Each sentence type condition consisted of 48 picture-sentence pairs; each picture was presented twice for counterbalancing of NP_1 and NP_2, and half of the pictures depicted the objects in the horizontal dimension and the other half in the vertical dimension (12 x 2 x 2 = 48).

The filler trials included 48 mismatching picture-sentence pairs and 96 matching picture-sentence pairs. Fillers were created by putting the preposition at the beginning of a sentence and by using the prepositions *naast* (next to) and *bij* (at).

Two experimental lists were created on the basis of these materials. Each list consisted of all the picture-sentence pairs divided over six blocks in such a way that a match, intra-dimensional mismatch and extra-dimensional mismatch of the same picture never occurred in the same block. List 1 and list 2 differed in the order of the blocks and were presented to an equal number of participants. Within each block the trials were pseudo-randomized with the following constraints: each block began with a filler trial, the same picture had to be separated by at least 3 other trials, a match or mismatch never occurred more than 3 times in a row, each sentence type condition never occurred more than 3 times in a row, an experimental or filler trial was never presented more than 3 times in a row, the dimension of the picture presented was never more than 3 times the same in a row, and the preposition used was never more than 3 times the same in a row.

2.3 Procedure

Participants were tested individually and seated in front of a computer screen in a dimly-lit Faraday cage. Both pictures and sentences were presented in black on a white background in a 10 cm by 10 cm window at the centre of the computer screen with a viewing distance of approximately 90 cm; making sure that participants could

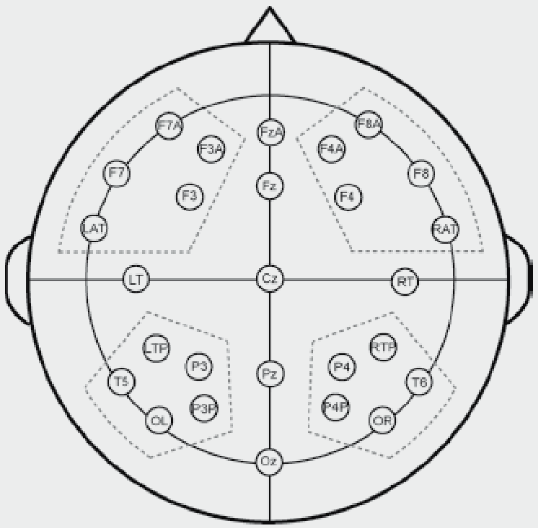
see the picture at one single glance. Presentation of the stimuli was accomplished by use of a Macintosh computer. A trial started with a fixation cross (duration 510 ms) followed by a 500 ms blank screen. Then a picture was presented for 2000 ms followed by a sentence in serial visual presentation mode. Word duration was 345 ms and the stimulus-onset asynchrony (SOA) was 645 ms. Sentence final words were indicated with a full stop. Inter-trial intervals lasted 1500 ms. Before starting the experiment, we explained our participants that the presented pictures would be one-dimensional. We showed them 3 example pictures and told them what the correct descriptions would be in terms of ‘in front of’-‘behind’ and ‘above’-‘below’. Furthermore, participants were instructed to attentively look at the picture and read the sentence. To avoid contamination of the EEG epochs of interest, participants were trained to postpone all eye movements until they had seen the sentence final word.

A reaction time (RT) pilot study was conducted, to test whether participants successfully detected the picture-sentence mismatches. A separate group of 15 participants was tested that fulfilled the same criteria as the participants in the EEG study. In the RT pilot study the participants had to press a button to indicate whether the sentence matched or mismatched the picture with their right (match) or left (mismatch) index finger.

2.4 EEG-recording

The electroencephalogram (EEG) was recorded with 27 tin electrodes mounted in an elastic electrode cap (Electrocap International; see Figure 1 for the montage). The electrode positions included standard International 10-20 system locations over the left and right hemispheres at the frontal (F3, F4, F7 and F8), midline (Fz, Cz, Pz and Oz), parietal (P3 and P4), and temporal (T5 and T6) sites. Eight extra electrodes were placed at the frontal (F3A, F4A, F7A and F8A), midline (Fza), and parietal (P3P and P4P) sites. In addition, eight electrodes were placed at nonstandard electrode positions previously found to be sensitive to language manipulations (e.g., Holcomb and Neville, 1990): left and right anteriortemporal sites (LAT and RAT: 50% of the distance between T3/4 and F7/8), left and right temporal sites (LT and RT: 33% of the interaural distance lateral to Cz), left and right temporoparietal sites (LTP and RTP: Wernicke’s area and its right hemisphere homologue: 30% of the interaural distance lateral to a point 13% of the nasion-inion

Figure 1 Electrode configuration used in the present experiment.



distance posterior to Cz), and left and right occipital sites (OL and OR: 50% of the distance between T5/6 and O1/2). Both the left and right mastoid were recorded; the right mastoid served as reference. Before the analyses, the signals were referenced to the average of the right and left mastoid. The electro-oculogram (EOG) was recorded bipolarly; vertical EOG was recorded by placing an electrode above and below the right eye and the horizontal EOG was recorded via a right to left canthal montage. A ground was placed on the forehead. For the EEG electrodes the impedance was always less than 3 kΩ, and for the EOG electrodes impedance was always less than 5kΩ. The signals were amplified (time constant = 8 s, bandpass = 0.02-30Hz) and digitized online at 200 Hz. Presentation of stimuli and recording of performance data were accomplished by a Macintosh computer.

2.5 Data analyses

Before analyzing EEG and EOG, records were examined for artifacts and for excessive EOG amplitude (>100 μV) from 100 ms before the onset of the critical preposition to 1000 ms following its onset. Averages were aligned to a 100-ms baseline period preceding the critical letter string; ending 1000 ms later. The ERPs

were analyzed in two ways. First, mean amplitudes were calculated in an early window (i.e., 200-400 ms) and a late window (i.e., 500-700 ms). These windows were based upon visual analysis and corresponded with the latency windows in which maximal differences between conditions were present.

Repeated measures analyses of variance (MANOVAs) were conducted separately for the midline sites and for the lateral sites with match (match, intra-dimensional mismatch, extra-dimensional mismatch) as factor. The multivariate approach to repeated measurements was used to avoid problems concerning sphericity (e.g., Vasey and Tayer, 1987). ERPs at the midline and lateral sites were analyzed in separate MANOVAs to examine laterality effects. The midline analysis included the additional factor site (Fza, Fz, Pz, Cz, Oz). To explore the scalp distribution of the ERP effects for the lateral analyses, we used a region of interest (ROI: anterior left: F3, F3a, F7a, F7, Lat vs anterior right: F4, F4a, F8a, F8, Rat vs. posterior left: P3, Ltp, T5, Ol, P3p vs posterior right: P4, Rtp, T6, Or, P4p; see also Figure 1) by hemisphere (left vs right) design. Interactions with the factor region of interest or hemisphere were followed up by additional analyses.

Another question was whether there are changes in ERP patterns over time. Because strategies may develop over time, a strategic modulation of reanalysis - as reflected by P600 - might be observed over different phases of the experiment. We tested this idea by comparing the ERP patterns during the early and the later stages of the experiment. The above described MANOVAs were conducted including the factor Block (first 24 items vs. last 24 items)

3 Results

3.1 Reaction time pilot study

MANOVAs were performed for the reaction times (RT) and error data with repeated measures on match (match vs. mismatch intra vs. mismatch extra).

For RT, a main effect of match ($F(2,14) = 9.92$, $p < .01$) revealed that mean RT for matches (1856 ms) was longer than for both the intra-dimensional mismatches

(1810. ms) and extra-dimensional mismatches (1803 ms; see Table 1). Follow-up analyses revealed that first; mean RT was longer for the matches in comparison to the intra-dimensional mismatches, ($F(1,15) = 12.34$, $p < .01$) and second; mean RT was longer for the matches in comparison to the extra-dimensional mismatches as well, ($F(1,15) = 17.88$, $p < .01$) and third; there was no significant difference in mean RT between the two kinds of mismatches, $F < 3.3$.

The error analyses revealed a main effect of match ($F(1,15) = 5.52$, $p < .05$). Follow-up analyses revealed that first; participants made significantly more errors on the matches than on the mismatches, ($F(1,15) = 11.81$, $p < .01$) and second; no other differences in error percentages between conditions were obtained, Fs<1.5 (see Table 1) [4]. There was no speed accuracy trade off.

Table 1 Mean reaction time (RT) in milliseconds and error percentages (Error) with standard deviations (S.D.), for the matches, intra-dimensional mismatches and extra-dimensional mismatches

Match				Intra mismatch				Extra mismatch			
RT	S.D.	Error	S.D.	RT	S.D.	Error	S.D.	RT	S.D.	Error	S.D.
1856.49	44.28	2.5	.004	1810.19	46.34	1.6	.006	1803.25	44.31	0.4	.003

[4] This pattern can appear unexpected since one might expect sentences describing the picture correctly to be processed faster and more accurate. It could be explained as follows: Because participants had to wait until they had read the last word of the sentence before they could know that a sentence described the depicted event correctly, they had to postpone their answer until the last word of the sentence was presented. However, participants could know that a sentence incorrectly described the picture as soon as the anomalous proposition was read. It is possible therefore, that in case of an incorrect sentence, the decision could be made earlier in the sentence and participants could already prepare the response during the sentence. Perhaps this led to faster mean reaction times and lower error percentages for both mismatching conditions.

2.2 Event-related potentials

Grand averages for the intra-dimensional mismatches and extra-dimensional mismatches time-locked to the onset of the critical propositions are presented in Figures 2 and 3, respectively. All conditions elicited an early ERP response characteristic for visual stimuli that is, an N1 which at occipital sites was preceded by a P1. Visual inspection of Figures 2 and 3 suggests that different patterns of brain activity were elicited by the match versus mismatch condition. For the match condition, an early positivity is followed by a negativity in the N400 window. In contrast, for the mismatch condition, an early negativity in the 200-400 ms time-window was followed by a late positivity which was most pronounced at central posterior sites. This biphasic brain response appears to be present for both the intra-dimensional and extra-dimensional mismatches. Furthermore, the difference waves presented in Figure 4 suggest that the time course of the early negative-going wave and the slow positive shift was very similar for the intra-dimensional and extra-dimensional mismatches.

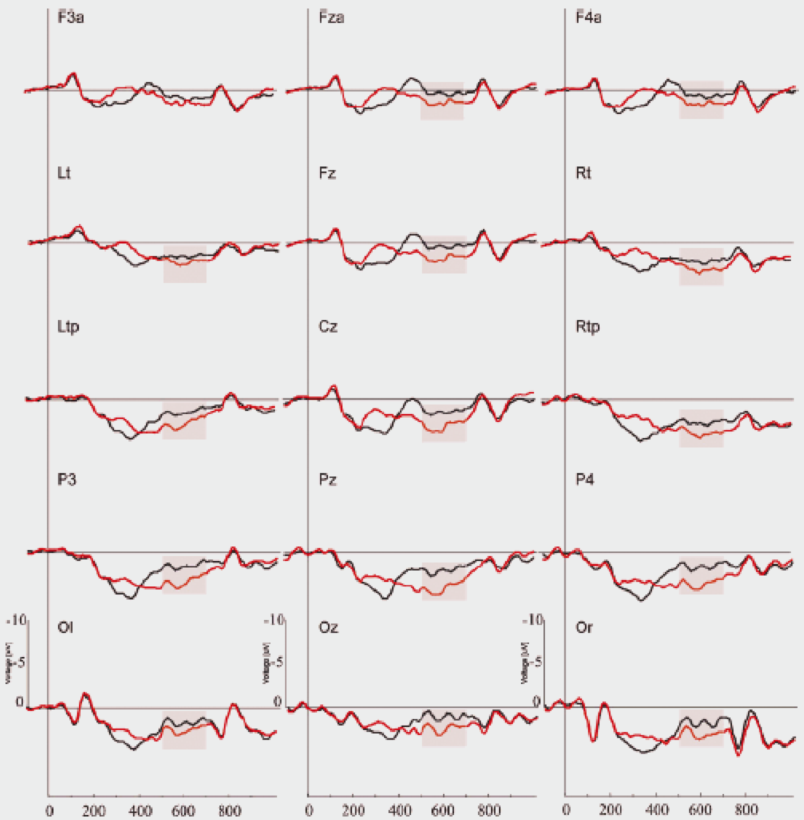
2.2.1. Statistical analyses

The mean percentage of trials that was rejected based on artifacts for the intra-dimensional, extra-dimensional mismatches and for the matches was 8.75%, 7.29%, and 7.02% respectively. As Table 2 shows, within the first 200 ms from target onset, no effects of match were found (all Fs < 1). Starting from 200 ms up to 800 ms main effects of match were obtained (all Fs > 5.61). Note that for the 200-300 ms and the 300-400 ms time windows, the effect of match reflects an increase in negativity for both the intra-dimensional and extra-dimensional mismatches. From 400 ms onward (400-500 ms, 500-600 ms, 600-700 ms, and 700-800 ms time window), the main effects of match unveils a different pattern, that is an increase in positivity for both the intra-dimensional and extra-dimensional mismatches. The report of the ERP results will be restricted to main effects and interactions that are relevant for the functional interpretation of the condition effects in the present study.

2.2.2 Early window (200-400)

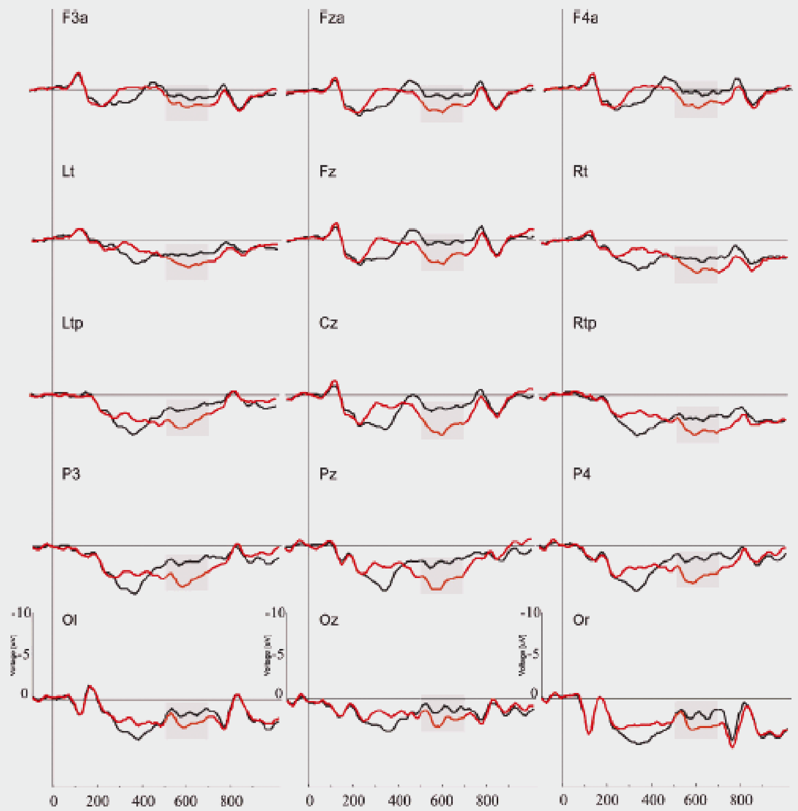
The omnibus analysis showed main effects of match for both the midline ($F(2,28) =$

Figure 2 Grand ERP averages for all midline and a subset of lateral sites, for matches (black line) versus intra-dimensional mismatches (red line). Averages are time-locked to the onset of the critical word, and superimposed for the two levels of match.



16.24, p <.001) and the lateral sites (F(2,28) = 22.86, p <.001). Follow-up analyses revealed that mean amplitude for the two kinds of mismatches was more negative-going compared to the match condition (midline and lateral: ps < .001). No difference in mean amplitude was found between the two kinds of mismatches (midline and lateral: Fs <1). In addition, the omnibus analyses for the lateral sites

Figure 3 Grand ERP averages for all midline and a subset of lateral sites, for matches (black line) versus extra-dimensional mismatches (red line). Averages are time-locked to the onset of the critical word, and superimposed for the two levels of match.



yielded an interaction of match by hemisphere (F(2,28) = 4.57, p <.02). Separate analyses for two hemispheres indicated that mean amplitude was more negative-going for the mismatches for both the left hemisphere (F(2,28) = 16.70, p <.001) and the right hemisphere (F(2,28) = 21.69, p <.001, see also Fig. 5 for the topographical maps).

2.2.3. Late window (500-700 ms)

The omnibus analysis revealed effects of match for both the midline sites (F(2,28) = 16.68, p <.001) and the lateral sites (F(2,28) = 11.91, p <.001). Follow up analyses indicated that overall mean amplitudes were more positive for the mismatches than for the matches (midline and lateral: ps <.001). In addition, there were no differences in mean amplitude between the intra-dimensional and extra-dimensional mismatches (midline and lateral: Fs<1.5). Furthermore, the omnibus analysis yielded a match by site interaction (F(8,22) = 7.24. p < .001) for the midline sites. Follow-up single site analyses revealed more positive amplitudes for the intra-dimensional mismatches and the extra-dimensional mismatches across the midline (ps < .05). For the lateral sites an interaction between match and region of interest was present (F(2,28) = 4.35, p <.05). Separate analyses for the anterior and posterior regions of interest indicated that for the posterior regions of interest mean amplitude was more positive for both kinds of mismatches than for the matches (F(2,28) = 19.53. p < .001), but that for the anterior regions of interest there were no differences in mean amplitude (F<3, see also Fig. 5 for the topographical maps).

Table 2 Presented are the F-values for the main effect of match for the midline analysis (Mid) and the lateral analysis (Lat) for the different 100 ms time-windows

Window	Match		Mismatch Intra		Mismatch Extra	
	F Mid	*F* Lat	*F* Mid	*F* Lat	*F* Mid	*F* Lat
0-100	<1	<1	<1	<1	<1	<1
100-200	<1	<1	<1	<1.5	<1	<1
200-300	17.97***	16.35***	10.52**	11.76**	18.03***	11.31**
300-400	28.29***	50.00***	14.12***	32.48***	39.02***	55.14***
400-500	16.48***	<2.5	21.57***	6.12*	5.61*	<1
500-600	40.56***	26.70***	23.84***	18.03***	38.88***	20.28***
600-700	24.14***	20.70***	15.45***	13.69**	23.62***	18.36***
700-800	6.66*	7.96**	<3.5	<3.5	8.05**	9.82**
800-900	<1	<1	<1	<1	<1	<1
900-1000	<3	<2.5	<2.5	<3.5	<2.5	<1.5

* p <.05, ** p <.01, *** < .001

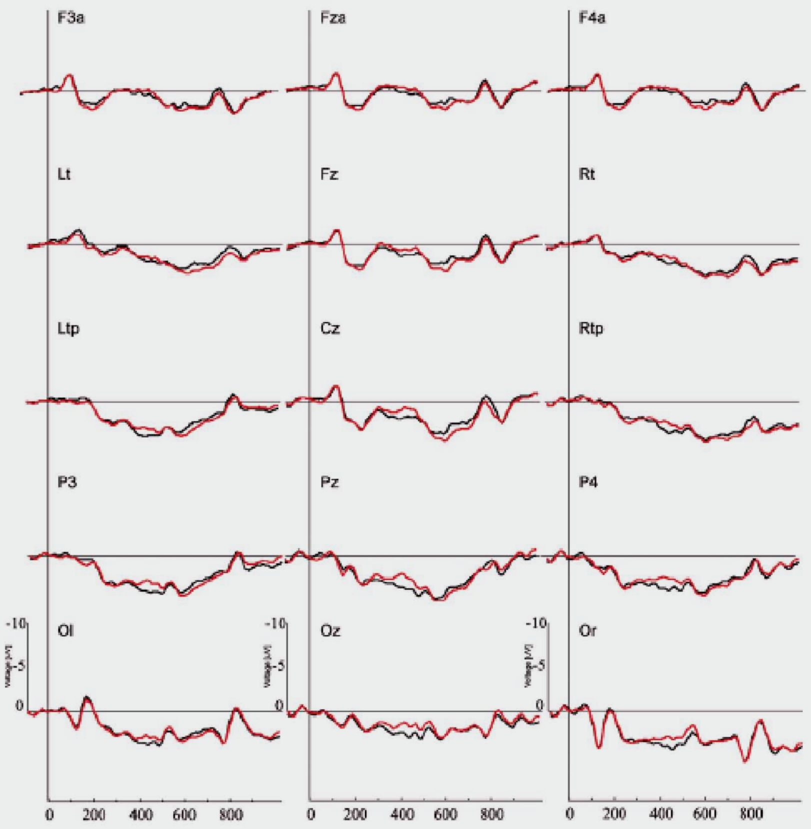
To test whether a modulation in the pattern of the late positivity was present over different phases of the experiment, we compared the late positivity during the early (first 24 items) and the later (last 24 items) block of the experiment. The omnibus analysis indicated that no main effect of block or interactions with block and match were obtained for the midline sites (Fs < 1). In addition, no main effects of block or interactions with block and match were present for the lateral sites (Fs < 1). These analyses revealed that the late positivity for the mismatches is equally present for both the early block and the late block. So, there are no indications for changes in the late positivity over time.

Discussion

The main findings of this article were as follows: a biphasic brain response was present for both the intra-dimensional mismatches and the extra-dimensional mismatches; that is, an early negative effect was followed by a late positive effect.

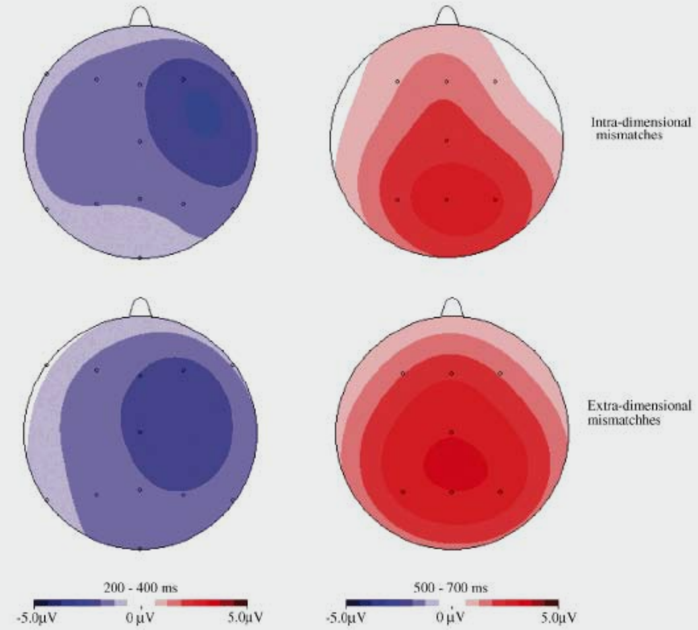
[5] We observed that the present P600 was centroparietally distributed across the scalp and that it extended to two anterior sites. One might argue that this could indicate that the P600 to picture-sentence mismatches is qualitatively different to the earlier reported P600. However, it must be realized that there is evidence for quite some variation in scalp distribution of the P600 after syntactic and semantic anomalies. A more frontal/broad distribution of the P600 effect has for example also been reported for locally ambiguous sentences (Friederici, Hahne, & Mecklinger, 1996; Hagoort, Brown, & Osterhout, 1999; Osterhout and Holcomb, 1995; Van Berkum et al., 1999) while Kaan and Swaab (2003) observed a more posterior distribution. As Figures 2 and 3 show, the present P600 had an early onset. With respect to P600 latencies, Friederici (1995) has proposed that differences in latency may reflect the complexity of processing necessary for the revision of the initially preferred reading. Longer latencies seem to be correlated with more complex restructuring. An early positivity, referred to as P345 has been observed after disambiguating auxiliaries (Friederici, 1995). Intermediate positivities have been shown to occur after syntactic and certain semantic anomalies (time range: 650-850 ms). Finally, even later positivities (time range: 700-1300 ms) have been reported in a recent discourse study (Nieuwland & Van Berkum, 2005). Hence, differences in latency may indeed be systematically related to the complexity of the language input that has to be checked for possible processing errors.

Figure 4 Difference waveforms for the intra-dimensional mismatch effect (intra-dimensional mismatches minus matches; black line) and the extra-dimensional mismatch effect (extra-dimensional mismatches minus matches, red line). The ERPs are time-locked to the onset of the critical word.



The late positivity resembled the P600 effect after syntactic and semantic violations both in timing and scalp distribution; that is, the effect showed a central/posterior scalp distribution and it started at 500 milliseconds and extended up to 800 milliseconds after the critical propositions (see also Kolk et al., 2003) [5]. So, both the intra-dimensional and extra-dimensional mismatches triggered a robust P600 effect.

Figure 5 Topographical maps obtained by interpolation from 27 sites for the two time-windows capturing the early negative effect (200-400) and the late positive effect (500-700). The first row shows the effects for intra-dimensional mismatches. The second row shows the effects for the extra-dimensional mismatches.



The block analyses indicated that there was no modulation in P600 pattern over different phases of the experiment. This finding can be taken to indicate that no strategies were developed over time during the picture-sentence matching task.

There are two possibilities to interpret the early negative effect. First, given the early latency of the effect, it could be an N2b; reflecting the mismatch between the conceptual representation based on the picture and the conceptual representation derived from the actual sentence. Wassenaar and Hagoort (2007) found a negativity with a similar distribution evoked by picture-sentence mismatches in the auditory domain. They interpreted the early negativity as an N2b and put forward that it reflects an on-line sensitivity to a picture-sentence mismatch. Likewise, D'Arcy, Connolly and Crocker (2000) observed an early negativity after a spoken sentence that incorrectly described a previously studied picture. They also labeled this negativity as an N2b and proposed it to reflect a deviation of the incoming spoken word with the active cognitive representation based on the visuospatial information.

A second explanation of the present negativity is that it reflects an N400-like effect. Van Petten and colleagues (1999) have shown that the onset of the N400 can be earlier when high semantic expectancy in sentential context is not confirmed. For the present study this could imply that the early onset of the negativity is caused by the violation of high semantic expectancy induced by the picture. After all, as described in the Introduction, we can assume that readers typically expect a sentence to be true. So, if a picture is presented, there will be a strong expectation to subsequently read a sentence, which describes the picture correctly. A sentence giving a false description, on the other hand, is highly unexpected. In addition, the broad topographical distribution of the negativity (present at both anterior and posterior sites) would be compatible with an interpretation in terms of N400.

Now the question is what could be the source of the P600 effect after picture-sentence mismatches in the present study? As described in the Introduction, P600 effects were observed after garden-path sentences (e.g, Osterhout & Holcomb, 1992) and also after different types of syntactic violations (e.g., Hagoort et al., 1993). However, the present intra-dimensional and extra-dimensional mismatches are syntactically unambiguous and do not contain syntactic anomalies either. So, syntactic ambiguity/violations can not be the trigger of the present P600 effect after intra-dimensional and extra-dimensional mismatches.

The P600 effect was observed to verb-argument semantic violations. As mentioned before, many of these semantic anomalies are repairable by reassigning the thematic roles of the arguments and ignoring the syntax (for a review see Kuperberg, 2007). On the base of this thematic reparability, the P600 effect after these semantic anomalies is proposed to reflect syntactic reanalysis; that is, to structurally repair a sentence by reassigning thematic roles (e.g., Kim & Osterhout, 2005; Kuperberg et al., 2003). As described in the Introduction, Wassenaar and Hagoort (2007) have also considered thematic role assignment as critical in eliciting a P600 effect after semantic verb-argument violations. They observed a P600 effect after sentences that mismatched the thematic roles displayed in a picture. They propose that the

role assignment on the basis of the picture interferes with the role assignment based on the sentence. Accordingly, the P600 effect was proposed to reflect how effortful the assignment process is.

The present P600 effect after picture-sentence mismatches cannot be related to thematic role (re)assignment as proposed by Kuperberg et al. (2003) nor to thematic role interference as proposed by Wassenaar and Hagoort (2007) because we ruled out potential thematic role assignment. The prepositions used in this experiment (e.g., *above, in front of*) do not implicate an action. Because thematic roles generally implicate an action, it is very unlikely that role assignment plays a role here. However, one may still think that location embodies a particular type of thematic role. One could accordingly assume that after an intra-dimensional mismatch like □ △ - *'the triangle stands in front of the square.'*, *'square'* and *'triangle'* have switched roles, and that this switching could be repaired by role assignment. As described in the Introduction, we added the extra-dimensional mismatches (e.g., □△ - '*The triangle stands below the square*') to definitively rule out a potential role assignment explanation. These mismatches are purely semantic and do not involve possible role switches; hence these mismatches can in no way be repaired by reassigning thematic roles.

As a supplement to thematic reparability, animacy constraints are proposed to be critical in eliciting a P600. Animacy is a grammatical-semantic property, which interacts in many ways with the process of thematic role assignment. Some verbs have an inherent thematic structure of Agent-Theme or Experiencer-Theme (Jackendoff, 1978). These verbs can only plausibly assign the role of Agent to an animate subject NP; an animacy violation is formed when these verbs are preceded by an inanimate subject NP. In a recent experiment by Packynski et al. (2006), such animacy violations evoked a P600 effect. They used sentences beginning with stems as: *'at long last the man's pain was understood by the* ….'. ERPs evoked by congruous animate nouns (e.g., *'doctor'*) were contrasted with incongruous animate nouns that were semantically associated (e.g., *'hypochondriac'*) or not associated (e.g., *'violinist'*) with the preceding context, and also with animacy violated inanimate nouns that were associated (e.g., *'medicine'*) or not associated (e.g., *'pens'*) with the preceding context. P600 effects were only present after animacy violations, regardless of semantic association. On the basis of this, Kuperberg (2007) proposes that animacy is a special semantic feature that has particular implications for determining thematic relationships between verbs and arguments. The present P600 effect cannot be explained by animacy violations, because the nouns used in this study are all inanimate.

Looking upon these data in the context of the current literature, a challenging question can be raised. What does the P600 effect evoked by syntactic violations (e.g., Hagoort et al., 1993), garden path sentences (e.g., Osterhout & Holcomb, 1992), semantic verb argument violations (e.g., Kolk et al., 2003), animacy violations (e.g., Packynski et al., 2006), orthographic violations in high cloze contexts (e.g., Vissers et al., 2006), picture-sentence mismatches (e.g., Wassenaar & Hagoort, 2007; present study) and most generally, after improbable events (Coulson, King, & Kutas, 1998a,b) have in common? The Monitoring Theory proposes a possible answer. It maintains that a sufficient trigger underlying the P600 effect is the presence of a conflict between two different representations of the same linguistic string. This conflict between representations triggers reprocessing to check whether the initial sentence processing has been correct. The P600 effect is proposed to reflect this process of reanalysis in language perception. For example, in garden path sentences, there is a conflict between the strongly preferred active interpretation on the basis of the syntactic bias and the less preferred correct interpretation in which the verb is taken as a passive particle. There is a similar mismatch between representations in semantic reversal anomalies. In these sentences the algorithmic parser and plausibility heuristic simultaneously propose incompatible thematic interpretations (e.g., *that poachers hunt foxes* vs. *that foxes hunt poachers*, see Kolk et al., 2003; Vissers et al., 2007). According to the Monitoring Theory, it is also a conflict between representations which triggers reanalysis after pseudohomophones in a high cloze context (Vissers et al, 2006). That is, a conflict between a highly expected noun on the basis of the high cloze context plus the phonological information which confirmed this expectation and made it maximally strong and the orthographic information.
Similarly, we are dealing with two strongly competing representations after the picture sentence mismatches of the present experiment. For both kinds of mismatches, there is a conflict between the conceptual representation based on the picture and the conceptual representation derived from the actual sentence. That is, the strong expectation emanating from the picture conflicts with the proposition which describes the picture incorrectly. The observed P600 effect is assumed to

reflect the immediate consequence of this conflict between the two conceptual representations.

There has been quite some debate about the functional relationship between the P600 and the P300 elicited by task-relevant, oddball stimuli. Some authors have claimed that the P600 is a P300-like effect (Coulson, King, & Kutas, 1998 a,b; Gunter et al., 1997; Van Herten et al., 2005). While other researchers have claimed that the P600 and the P300 are distinct components (Osterhout & Hagoort, 1999; Frisch et al., 2003). Whether the late positivity in the present experiment is a P600 or a P300 remains an empirical question. The monitoring theory proposes that the P600 elicited by linguistic violations could very well be a variant of the P300 (see also Van Herten et al, 2005). Both P600 and P300 are sensitive to probability manipulations, with larger positivities to less probable events (Coulson et al., 1998 a, b; Hahne & Friederici, 1999; Gunter et al., 1997, Vissers et al., 2007). In the present study, the positivity to both kinds of mismatches could reflect the unexpectedness of the prepositions in the mismatching conditions as opposed to the expectedness of the prepositions in the matching condition. The monitoring hypothesis suggests a specific reason for this sensitivity to expectancy: the less an event is expected, the more chance there is that the event is due to erroneous processing. Furthermore, the fact that the P300 and the P600 have a very similar centro-parietal scalp distribution let some authors to conclude that the two components are identical. Latency differences between the P600 and the P300 oddball effect are explained with the higher complexity of linguistic stimuli. The monitoring theory proposes that the large differences in the timing of the P300 and P600 (from about 300 up to 1100 post stimulus) can be seen as a function of the type and complexity of the unexpected material that has to be reprocessed [6]. Just as the latency of the P300 has been shown to be a function of the stimulus evaluation time (e.g., Donchin, 1979), the latency of the late positivity may vary with the difficulty of checking the perceptual input or its memory trace for possible processing errors.

[6] Longer latencies seem to be correlated with more complex restructuring. An early positivity, referred to as P345 has been observed after disambiguating auxiliaries (Friederici, 1995). Intermediate positivities have been shown to occur after syntactic and certain semantic anomalies (time range: 650-850 ms). Finally, even later positivities (time range: 700-1300 ms) have been reported in a recent discourse study (Nieuwland & Van Berkum, 2005).

In conclusion we can say that the P600 does reflect reprocessing. The function of this reprocessing is a general one, to find out whether one has read the sentence correctly. This general check for processing errors can occur after syntactic violations, garden path sentences, semantic verb argument violations, animacy violations, orthographic violations in high cloze contexts, picture-sentence mismatches and most generally, after improbable events.

Acknowledgments

Portions of this research were presented at the 14th Annual Meeting of Cognitive Neuroscience Society in New York, 2007. We thank two anonymous reviewers for their very helpful comments on an earlier version of this article. We are grateful to the ERG group of the NICI for technical assistance and to Maarten vd Meulen for his support.

References

D'Arcy, R.C.N., Connolly, J.F., & Crocker, S.F. (2000). Latency shifts in the N2b component track phonological deviations in spoken words. *Clinical Neurophysiology*, 111, 40-44.

Chwilla, D.J., Brown, C.M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32, 274-285.

Chwilla, D.J., Hagoort, P., & Brown, C.M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. The Quarterly *Journal of Experimental Psychology A: Human Experimental Psychology*, 51A, 531-560.

Coulson, S., King, J., & Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, 13, 653-672.

Coulson, S., King, J., & Kutas, M. (1998b). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes,* 13, 21-58.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begeleiter (Ed.), *Evoked Potentials and Behavior* (pp. 13-75). New York: Plenum.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.

Friederici, A.D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language,* 50, 259-281.

Friederici, A.D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219-1248.

Friederici, A.D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, 25, 157-176.

Friederici, A.D., Pfeifer, E. & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1, 183-192.

Friederici, A.D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition*, 27, 438-453.

Frisch, A., Kotz, S.A., Von Cramon, D.Y., Friederici, A.D. (2003). Why the P600 is not just a P300: the role of the basal ganglia. *Clinical Neurophysiology*, 114, 336-340.

Gunter, T.C., Stowe, L.A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.

Hagoort, P., Brown, C.M, & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.

Hagoort, P., Brown, C.M., & Osterhout, L. (1999) The neurocognition of syntactic processing. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273-317). Oxford: University Press.

Hahne, A., & Friederici, A.D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194-205.

Hoeks, J.C.J., Stowe, L.A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59-73.

Holcomb, P.J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30, 47-61.

Holcomb, P.J., & Neville, H.J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, 5, 281-312.

Jackendoff, R. (1978). Grammar as evidence for conceptual structure. In M. Halle, J. Bresnan & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 201-228). MIT press, Cambridge, MA,.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.

Kaan, E., & Swaab, T.Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience* 15, 98-110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1-36.

Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.

Kuperberg, G.R., Kreher, D.A., Sitnikova, T., Caplan, D.N., & Holcomb, P.J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100, 223-237.

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117-129.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.

Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.

Neville, H.J., Nicol, J.L., Barss, A., Forster, K.I., & Garett, M.F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151-165.

Nieuwland, M.S., & Van Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24, 691-701.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, 14, 1-14.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Osterhout, L., & Holcomb, P. J. (1995). Event-related brain potentials and language comprehension. In M.D. Rugg & M.G.H. Coles (Eds.), Electrophysiology of mind: *Event-related brain potentials and cognition* (pp171-215). New York: Oxford University Press.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 786-803.

Osterhout, L., & Mobley, L.A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739-773.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283-317.

Packynski, M., Kreher, D.A., Ditman, T., Holcomb, P.J., & Kuperberg, G.R. (2006) Electrophysiological evidence for the role of animacy and lexico-semantic associations in processing nouns within passive structures. *Cognitive Neuroscience Supplements*, Abstract

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447-481.

St. George, M., Mannes, S., & Hoffman, J.E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6, 70-83.

Van Herten, M., Chwilla, D.J., & Kolk, H.J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitve Neuroscience*, 18, 1181-1197.

Van Herten, M., Kolk, H.H.J., & Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word-frequency in event-related brain potentials. *Memory and Cognition*, 18, 380-393.

Van Petten, C., Coulson, S., Rubin, S., Plante, S., & Parks, M. (1999), Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 394-417.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.

Vasey, M.W., & Thayer, J.F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, 24, 479-486.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18.

Vos, S.H., Gunter, T.C., Kolk, H.H.J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41-63.

Wassenaar, M., & Hagoort, P. (2007) Thematic role assignment in patients with Broca's aphasia: Sentence-picture matching electrified. *Neuropsychologia*, 45, 716-740.

Chapter 5

Summary and Conclusions

Chapter 5: Summary and Conclusions

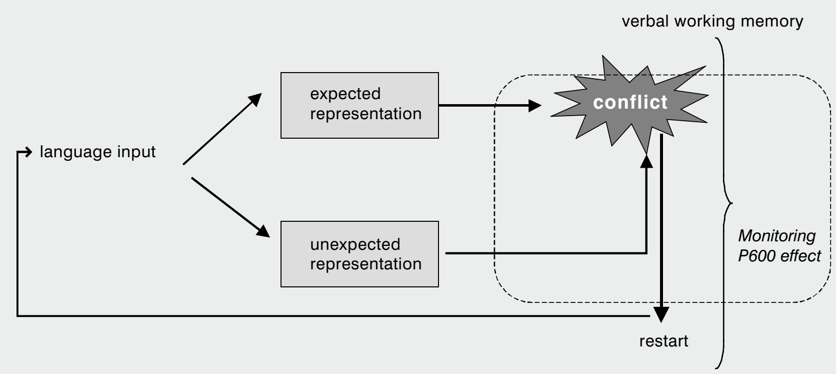
In this chapter the major findings of the experimental chapters are summarized and a schematic representation is given of how according to the Monitoring Theory perceptual errors are monitored for. Finally, frequently asked questions with respect to the Monitoring Theory will be addressed and suggestions for future research will be discussed.

Monitoring for perceptual errors

As described in the General Introduction, although error monitoring has been exclusively studied in production, we also make perceptual errors, for example when we misunderstand a speaker or when we misread a word. In the case of perceived speech, errors can not be observed directly in the way as they can in produced speech. So a crucial question addressed in this thesis is how could the language user know that a perceptual error occurred? This is the first project that systematically tests for monitoring in the domain of language perception and attempts to explain how monitoring for perceptual errors comes about. In the present thesis, we propose that it is a strong conflict between what is expected and what is observed which signals the presence of a possible processing error. Thereby, a monitoring response is triggered to check for possible perceptual errors. We maintain that it is this process which is reflected by the P600. The results of the experimental chapters demonstrate that monitoring takes places at different linguistic levels (sentence level, word level and conceptual level). From this we conclude that monitoring plays an important role in language perception.

Let us now explain in more detail how according to our theory perceptual errors can be monitored for by the reader. A schematic representation of how possible errors in language perception are detected and filtered out is depicted in Figure 1. The simultaneous activation of an expected representation and an unexpected representation of the same linguistic string leads to a conflict. This conflict is proposed to bring the reader to take seriously the possibility that he has misread the sentence. Therefore the reader starts reprocessing the input to check whether it has been perceived appropriately. The P600 effect is proposed to reflect the

Figure 1 Schematic representation of how the language system deals with the consequence of a possible processing error, according to the Monitoring Theory.



reanalysis to check for possible processing errors. The verbal working memory is the arena where conflicts are resolved.

We propose that the P600 effect reflects an aspect of cognitive or executive control of language. Executive control is necessary in the case of response uncertainty; that is, whenever there is more than one way to respond in a particular situation. Some recent ERP studies provide clear evidence for top-down influences; namely that language users online generate expectancies for upcoming words (e.g., Delong et al., 2005; Van Berkum, Hagoort, & Brown, 1999). For instance, in some cases it is possible to predict from the context a particular word, this can make this word easy to recognize. In the sentence *'In that library the pupils borrow books.'* the word *'books'* is almost 100% predictable and is therefore easily recognized. This is an example of a top-down effect without response uncertainty. In case of response uncertainty, there is a predictive context, but this context allows for several possibilities. It is commonly assumed that the attentional system is needed to select one of the alternatives. We propose that top-down processes play an important role in language processing; that is, a process in need of executive control. The P600 effect is proposed to reflect one aspect of cognitive control, namely monitoring for errors.

A general check for processing errors can occur for example after garden-path sentences, which are known to reliably elicit a P600 effect. Readers of a sentence like '*the woman persuaded to answer the door.*', chose first for an active interpretation of the verb persuade and thus assume that the sentence is about a woman persuading someone. Consequently, they expect an object NP after the verb, but read *to* instead. So, in garden-path sentences, the active interpretation which is strongly *expected* on the basis of the syntactic bias conflicts with the *unexpected* correct interpretation in which the verb is taken as a passive particle. This conflict between representations forces the reader to reanalyze the sentence, in order to arrive at an interpretation in which the verb is taken as a passive participle and thus realize that the sentence is about a woman being persuaded.

Likewise the Monitoring Theory proposes that the P600 effect after semantic anomalies reflects reanalysis triggered by a conflict between two incompatible representations. In particular, when reading a sentence like *'the fox that hunted the poachers.',* plausibility heuristics and syntactic algorithms produce different thematic interpretations. Whereas the plausibility heuristic leads to the most plausible, *expected* interpretation, that is that 'the poachers are hunting the foxes', parsing routines lead to the implausible, *unexpected* interpretation that *'the foxes are hunting the poachers'*. In both garden-path sentences and semantic reversal anomalies it makes sense for the reader to check the correctness of his or her analysis. Or to wonder 'did I read that correctly?'. After all, an inconsistency can have two sources. It can be real, in the sense that an unexpected event did indeed occur (that the man bites the dog). On the other hand it could stem from a processing error. Therefore, filtering out errors is essential for proper language comprehension.

The P600 effect is proposed to reflect reprocessing triggered by the activation of two incompatible representations. This reprocessing has the more *general function* of checking for possible processing errors. Because different linguistic elements can be misperceived, the P600 was predicted to reflect reprocessing at different linguistic levels. The results of the experiments described in Chapter 2, 3 and 4 support our Monitoring Theory in that they show that the P600 effect can reflect sentence reanalysis at different linguistic levels. That is; monitoring at the sentence level, at the word level and at the conceptual level.

Monitoring at the sentence level

In the experiment described in Chapter 2 (Vissers, Chwilla, & Kolk, 2007) we test the hypothesis that, as described above, the P600 effect after semantic anomalies like *'The fox that hunted the poacher.'* reflects monitoring triggered by a conflict between the outcome of a lexical strategy with that of the parsing routine ((Kolk, Chwilla, van Herten, & Oor, 2003). Whereas the plausibility heuristic leads to the interpretation that *the poachers hunted the foxes*, the parsing routines lead to the interpretation that *the foxes hunted the poachers*. Kolk and colleagues propose that this conflict between the semantically plausible, highly expected thematic interpretation and the implausible thematic interpretation makes it necessary for the brain to re-attend the unexpected linguistic unit to check its veridicality. The P600 effect observed after these semantic reversal anomalies is assumed to reflect this general check for processing errors at the sentence level.

We test this hypothesis by using the same stimulus materials as in the study by Kolk and colleagues while manipulating the instruction. In the Kolk and colleagues' study (2003), participants had to indicate if the sentence was semantically acceptable or not. In contrast, in the study reported in Chapter 2 participants are told that semantic reversals had been created on purpose. Specifically they are told that they should not be misled by their knowledge of what normally happens in the world, but pay extra attention to "who does what to whom". The rational for this instruction is that it should reduce the discrepancy between the thematic interpretations proposed by the semantic heuristic and the syntactic parse. If Kolk and colleagues' proposal (2003) that the P600 effect after semantic reversal anomalies is based on a control operation triggered by a conflict at the sentence level is correct, then this focus-on-syntax instruction should lead to a decrease in error rates and a reduction or disappearance of the P600 effect.

Our focus-on-syntax instruction does influence both the behavioural data and ERP data as predicted. First, participants are faster and more accurate than the participants in the Kolk et al. (2003) study. This improvement in performance can be taken to indicate that the focus-on-syntax instruction is effective, in that our participants are less easily misled by the semantic reversals. Second, the focus-on-syntax instruction leads to a disappearance of the P600 effect at the midline sites and at all but one site of the left hemisphere. Because the instruction turns an unexpected real life event (that foxes hunt poachers) into a less unexpected event, there is less need for the brain to re-attend the implausible linguistic unit which resulted in a decrement of the P600 effect. The results of the present study support the proposal by Kolk and colleagues (2003) that the P600 effect to semantic reversal anomalies is based on a control operation triggered by a conflict between two incompatible interpretations at the sentence level; that is a conflict between an expected thematic interpretation proposed by the plausibility heuristic and an unexpected thematic representation proposed by the algorithmic parser.

Monitoring at the word level

In the experiment described in Chapter 3 (Vissers, Chwilla, & Kolk, 2006) we investigate whether a conflict between a highly expected representation and an unexpected representation can also trigger a monitoring response at the word level. To this aim, we create both a high and a low-cloze context for a particular lexical item. There is no difference in high and low cloze context sentences except for the critical item ('The pillows are stuffed with *feathers* which make them feel soft.' versus 'The pillows are stuffed with *books* which makes them feel hard'). The critical lexical item is either spelled correctly or is a pseudohomophone derived from the expected word and phonetically similar to the expected word.

We predict that the pseudohomophone/high-cloze sentences would lead to a conflict at the word level, between the tendency to accept the pseudohomophone, and the tendency to reject it. The tendency to accept the pseudohomophone is supposed to be very strong because it corresponds to a word which was semantically highly expected and also because the phonological form of the word confirmed this expectation and makes it maximally strong. On the other hand, the tendency to reject the pseudohomophone will also be very strong, because it is misspelled. The resulting conflict is expected to bring the brain into a state of indecision and elicit a monitoring response. We predict that this monitoring process will be reflected by a P600 effect. Second, for the low cloze sentences, the lexical items from which the pseudohomophones are derived are not highly expected and

thus should not elicit a conflict between the expected and actually presented lexical item. Consequently, no monitoring process and hence no P600 effect is expected to occur.

The ERP data confirm the prediction in that only pseudohomophones embedded in a high-cloze context gave rise to a P600 effect. Because the words from which the pseudohomophones are derived are highly expected, initially the pseudohomophones are easily integrated into the higher order meaning representation of the context. After all, the phonological representation of the pseudohomophone is congruent with the sentential constraints. But when the subject detects the misspelling, which signals a possible processing error, a monitoring response is triggered. This monitoring process is reflected by the P600 effect. In the low cloze sentences, the words and pseudohomophones are not expected. Consequently, the pseudohomophones are not qualified as possible processing errors and do not confuse the reader; hence, no monitoring process or P600 is triggered. As predicted, a P600 effect is only elicited in cases of a strong conflict, when an unexpected linguistic event is observed while another event is predicted with more or less 100 % certainty.

The results of Chapter 3 support the view that in addition to a monitoring process at the sentence level, there is a process of monitoring in language perception at the word level. It occurs whenever a conflict between a strong tendency to accept and one to reject a word brings the cognitive system in state of indecision. To resolve this conflict between two incompatible representations, the brain monitors the input to check for possible processing errors. This monitoring response at the word level is reflected by the P600.

Monitoring at the conceptual level

In the experiment described in Chapter 4 (Vissers, Chwilla, Van de Meerendonk, & Kolk, 2008) we investigate whether a conflict between a highly expected representation and an unexpected representation can also trigger a monitoring response at the conceptual level after picture-sentence mismatches. To test this, expectancy is manipulated by varying the veridicality value between a picture and a sentence. Wassenaar and Hagoort (2007) observe a P600 effect after sentences that contain thematic roles that were incompatible with the thematic information perceived from a previously presented picture. For example after a picture in which a woman pushes a man in a wheelchair, the sentence *'the tall man on this picture pushes the young woman'* gives rise to a P600. Wassenaar and Hagoort propose that in these sentences the role assignment based on the picture interferes with the role assignment based on the sentence; since the picture indicates one role assignment, it is difficult to assign the reverse thematic roles in the sentence. The size of the P600 effect is suggested to vary as a function of how effortful the assignment process is. The Monitoring Theory accounts for the P600 effect in a different way, by proposing that it reflects reanalysis for monitoring purposes triggered by the conflict between the predicted thematic roles on the basis of the picture representation and the thematic roles based on the sentence. In the present study, we exclude thematic roles, to circumvent the notion that any P600 effect is attributed exclusively to reanalysis for syntactic purposes; that is, reanalysis to structurally repair a sentence (e.g., Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Wassenaar & Hagoort, 2007)

To this aim, we present pictures of spatial arrays, followed by a sentence giving a correct or an incorrect description of the picture, depending on the preposition being used. We use two different types of mismatch; namely, intra-dimensional mismatches (e.g, □ △ - 'De driehoek staat *voor* het vierkant.' Paraphrase: 'The triangle stands *in front of* the square.'), and extra-dimensional mismatches (e.g., □△ - 'De driehoek staat *boven* het vierkant.' Paraphrase: 'The triangle stands *above* the square.'). By employing locative relationships, we avoid the involvement of thematic role assignment and thereby a possible effect of role assignment interference, as proposed by Wassenaar and Hagoort (2007). This is because locative relationships do not implicate an action while thematic roles generally do implicate an action. Therefore, it is very unlikely that role assignment plays a role here. However, one may still argue that location embodies a particular type of thematic role. Accordingly, one could presume that after an intra-dimensional mismatch like □ △ - *'the triangle stands in front of the square.'*, *'square'* and *'triangle'* have switched roles, and that this switching could be repaired by role reassignment. We add the extra-dimensional mismatches (e.g., □ △ - *The triangle stands below the square*) to definitively rule out a potential role assignment explanation. These mismatches are purely semantic and do not involve possible

role switches; hence reassigning thematic roles would not be helpful in this condition. So any P600 effect observed under these circumstances cannot be interpreted as reflecting reanalysis for syntactic purposes.

As predicted, a robust P600 effect is present after both intra-dimensional mismatches and extra-dimensional mismatches. We propose that the trigger underlying this P600 effect is the presence of a conflict between the conceptual representation based on the picture and the conceptual representation derived from the actual sentence. Specifically, the strong expectation emanating from the picture conflicts with the proposition which describes the sentence incorrectly. We assume that this conflict between conceptual representations triggers reanalysis reflected by the P600 effect. As described above, because we ruled out potential thematic role assignment the present P600 effect cannot be related to reanalysis for syntactic purposes. We propose that the P600 effect reflects reprocessing for monitoring purposes, to find out whether one has read the sentence correctly.

Conclusions

In summary, the experimental findings presented in this thesis yield a consistent picture with regard to the question whether monitoring occurs in language perception. The main assumption underlying this research project is that the P600 reflects a more general process of reanalysis to check and filter out potential processing errors. Based on this assumption we predict that since different linguistic elements can be misperceived, a P600 effect should be elicited at different linguistic levels. Each of the three experimental chapters supports this assumption:

In Chapter 2 we show that the P600 is strongly modulated by instruction which indicates that a substantial part of the P600 effect can be accounted for by a control operation triggered by a conflict at the sentence level. In Chapter 3 we show that a conflict at the word level yields a P600 effect: In particular, we demonstrate that pseudohomophones produce a P600 effect when embedded in a high-cloze context but not when embedded in a low-cloze context. This reveals that a P600 effect is only elicited after a strong conflict between the tendency to accept and the tendency to reject a particular word. In Chapter 4 we test the monitoring hypothesis by inducing a strong expectation in a very different way, namely by varying the veridicality/truth value of the relation between a picture and a sentence. In line with the Monitoring Theory a P600 effect was elicited by a conflict between a highly expected conceptual representation based on the picture and the representation of the sentence.

Taken together, the present ERP results provide strong evidence for the claim that monitoring indeed does occur in language perception. How does the monitoring hypothesis relate to other reanalysis accounts that have been proposed in the literature? The main difference between the monitoring hypothesis and other reanalysis accounts of the P600 concerns the function of the reanalysis. Other accounts claim that reanalysis occurs only for syntactic purposes while the monitoring hypothesis claims that the reanalysis has a more general function –that is, to check and filter out processing errors to prevent integration of erroneous information. At a more general level the monitoring process in language perception and language production enables readers and speakers to understand language in the fast and efficient way they do. We have shown that event-related brain potentials are an ideal method to catch the dynamics of this highly complex cognitive function.

Frequently asked questions

In the process of presenting my findings over the passed four years a number of important questions have been posed by colleagues in the field. In this section I address each of these questions and detail the relevant aspects of the Monitoring Theory based on my findings presented in this thesis.

Does monitoring detect a conflict or is monitoring triggered by a conflict?

The first question concerns the conflict that lies at the heart of the monitoring process. Is it the conflict triggering a monitoring response or is it the monitoring response detecting a conflict? In the present dissertation I propose that it is the conflict which triggers a monitoring response which elicits a P600 effect rather than the other way around. That is, we assume that a conflict needs no monitoring to be detected. It is detected automatically because it brings the system into a state of indecision. The P600 is triggered when a conflict evolves because the brain

encounters an unexpected linguistic item when another item is highly expected. To resolve this conflict between representations, the brain reprocesses the input to check for possible processing errors. I will illustrate this point on the basis of the data of Chapter 3 in which I test for monitoring at the word level. In this chapter I propose that a *conflict* triggers a monitoring process in perception, similar to what has been shown in the action domain (for a further discussion see Van Herten, Chwilla, & Kolk, 2006). Specifically, I propose that the conflict arises between different tendencies: the tendency to reject and the tendency to accept the word. The prediction of this study is that the pseudohomophone/high-cloze sentences would lead to a conflict at the word level, between the tendency to accept the pseudohomophone, and the tendency to reject it. The tendency to accept the pseudohomophone is supposed to be very strong not only because it corresponds to a word which is semantically highly expected but also because the phonological form of the word confirms this expectation and makes it maximally strong. On the other hand, the tendency to reject the pseudohomophone would also be very strong, because it is orthographically ill-formed. The resulting conflict is expected to bring the brain into a state of indecision and elicit a monitoring response that should elicit a P600. In the low cloze condition, the lexical items from which the pseudohomophones are derived are not expected and thus should not create a mismatch between an expected and an actually presented lexical item. Hence, no monitoring process and no P600 are expected to occur.

Why isn't the P600 effect elicited by any unexpected event?

One may wonder why we do not predict a P600 effect after any unexpected event. And indeed, why does a typical N400 sentence like *'He spread the warm bread with socks'* (Kutas & Hillyard, 1980) not elicit a P600? After all, in these sentences there is also a conflict between an expected and an unexpected linguistic event, which could trigger a monitoring response and hence a P600.

As argued above, the function of the monitoring process is to edit out possible processing errors. Such a process should not be triggered by every single unexpected linguistic unit. Informative statements are always somewhat unexpected but they should be normally integrated into the discourse information. Checking for possible processing errors too often would be disruptive for communication.

We thus assume that only strong conflicts will trigger a monitoring response. Mild conflicts such as ones created by a relatively expected and a relatively unexpected word will not elicit a monitoring response. For instance, in the case of misspellings in a high-cloze context (Vissers et al., 2006), there is a strong conflict between a representation supported by both semantic and phonological information and a representation delivered by bottom up orthographic analysis; this strong conflict brings participants to take seriously the possibility that they have misread the sentence. In the low cloze sentences, the pseudohomophones are not as expected as in the high-cloze sentences and thus do not make the brain to disbelieve what was read. Therefore, the pseudohomophones in the low cloze sentences give rise to an N400 effect of cloze probability (and not to a P600 effect). A P600 effect is only elicited in cases of a strong conflict, when an unexpected linguistic event is observed while another event is predicted with more or less 100% certainty. In other words, monitoring will not set in after encountering any anomaly, but only when there is a high degree of uncertainty about the source of the anomaly: was it really there or could it stem from a processing error?

In a recent study in our lab, the hypothesis that the degree of unexpectedness determines whether an N400 effect or a P600 effect is triggered is further explored (Van de Meerendonk, Vissers, Chwilla, & Kolk, submitted). To this aim, three experimental conditions are created: plausible sentences: e.g., '*The eye consisting of among others things a pupil, iris and retina* …', mildly implausible sentences: e.g., '*The eye consisting of among others things a pupil, iris and eyebrow* …' and deeply implausible sentences: e.g., '*The eye consisting of among others things a pupil, iris and sticker* …'. The mildly implausible sentence condition is hypothesized to trigger a mild conflict between the expected and the unexpected critical noun. As predicted, the mild unexpectedness leads to integration difficulties which are reflected in an N400 effect for this condition. For the deeply implausible sentences, we hypothesize a strong conflict to occur between the expected and the unexpected noun. As predicted, these highly implausible sentences cause integration failure *and* a monitoring process which is reflected in a biphasic N400-P600 pattern. Because these sentences are completely unexpected, one cannot integrate this information into a knowledge base (reflected by an N400 effect). This integration failure brings the brain to mistrust what was read, therefore a monitoring response is triggered (reflected by a P600 effect); 'Did I read that correctly?'.

What is it that is reprocessed?

As pointed out and illustrated with several examples from speech production (slips of the tongue) and speech perception (slips of the ear) in the Introduction of this dissertation people filter out possible perceptual errors. According to the Monitoring Theory it is the input in all its aspects (including phonology, orthography, semantics and syntax) that is processed again to prevent integrating erroneous information. This process is similar to what often happens in daily conversation. If a listener receives unexpected input from an interlocutor (e.g., "that a man bites a dog" or "that planes crashed into the World Trade Center in New York") people ask whether what they heard was correct. This checking process optimizes the comprehension process by attempting to prevent perceptual errors to be integrated into the ongoing discourse.

Can we still speak of one P600 effect given the observed variability in scalp distribution?

In previous studies, the P600 effect to semantic and syntactic violations shows a central/posterior scalp distribution (Coulson et al., 1998; Kolk et al., 2003). The P600 effect observed after misspellings (Chapter 3) is centroparietally distributed across the scalp extending to two right anterior sites. In addition, the P600 effect after picture-sentence mismatches (Chapter 4) is broadly distributed and also present at anterior sites. One of the guidelines for interpreting ERP waveforms (see for example Kutas, 1993) is that differences in waveshape and/or scalp distribution between two or more conditions reflect the activity of distinct neuronal populations subserving qualitatively different processes. Therefore, one might argue that differences in the distribution of the P600 effect (e.g. to conflicts at the word level or the conceptual level) are qualitatively different from the standard syntactic P600 effect.

I would like to point out, however, that there is evidence in the literature for quite some variation in scalp distribution of the P600, even after syntactic anomalies which are the classical case for the occurrence of the P600. In particular, a more frontal and broad distribution has been reported for locally ambiguous sentences by Friederici and colleagues (1996) and also by Osterhout and Holcomb (1995). While a more posterior distribution was observed by Kaan and Swaab (2003).

Based on this variability in the literature on P600 effects to syntactic anomalies, I do not think that the reported variability in scalp distribution in the present studies has to be attributed to qualitatively different processes. According to the Monitoring Theory, one would allow that the P600 has some variation in scalp distribution given the fact that different levels of (re) processing are involved. In other words, the type of material that has to be processed (a single word versus a whole clause), could very well affect the scalp distribution of the P600 effect. A definite answer to this question requires a direct within subject comparison elicited by semantic, syntactic and conceptual P600 effects.

Is the P600 qualitatively different from the P300?

There has been quite some debate about the functional relationship between the P600 and the P300 elicited by task-relevant, oddball stimuli. Some authors have claimed that the P600 is a P300-like effect (e.g., Coulson, King, & Kutas, 1998; Gunter et al., 1997), while other researchers have claimed that the P600 and the P300 are distinct components (Osterhout & Hagoort, 1999; Frisch et al., 2003). Applied to Chapter 4, one could ask whether the late positivity to conceptual anomalies is a P600 or a P300. My reply to this is that this remains an empirical question. The monitoring theory proposes that the P600 elicited by linguistic violations could very well be a variant of the P300 (see also Van Herten et al., 2006). Both P600 and P300 are sensitive to probability manipulations, with larger positivities to less probable events (Coulson et al., 1998; Hahne & Friederici, 1999; Gunter et al., 1997, Vissers et al., 2007). Therefore the positivity in chapter 4 to intra-dimensional and extra-dimensional mismatches could reflect the unexpectedness of the prepositions in the mismatching conditions as opposed to the expectedness of the prepositions in the matching condition. The monitoring hypothesis suggests a specific reason for this sensitivity to expectancy: the less an event is expected, the more chance there is that the event is due to erroneous processing. Furthermore, the fact that the P300 and the P600 have a very similar centro-parietal scalp distribution let some authors to conclude that the two components are identical. Latency differences between the P600 and the P300 oddball effect are explained with the higher complexity of linguistic stimuli. The Monitoring Theory proposes that the large variation in the timing of both the P300 and the P600 (from about 300 up to 1100 post stimulus) can be seen as a function of the type and complexity of the

unexpected material that has to be reprocessed. Just as the latency of the P300 has been shown to be a function of the stimulus evaluation time (e.g., Donchin, 1979), the latency of the late positivity may vary with the difficulty of checking the perceptual input or its memory trace for possible processing errors.

Are you investigating perceptual errors as such or how the language system deals with the consequences of a perceptual error?

With our stimulus materials, we simulate the consequences of a perceptual error. We investigate how the language system deals with perceptual errors, but we do not study the genesis of perceptual errors. In other words, we investigate the process from the moment that there is an unexpected event in the language comprehension system, and we do not study the process which leads to the unexpected event.

There will be differences between the errors in the stimulus materials of this dissertation and the spontaneous misperceptions we make in daily life (e.g., 'slips of the ear'). The perceptual errors that arise in ordinary conversations have specific syntactic, semantic and phonological properties that we can only try to mimic in our experiments. However, both in ordinary conversations and in our experiments, errors result in abnormal representations. For the language system, there is no difference between the abnormal representation after misperceptions in daily life and the abnormal representation triggered by the stimulus materials of the present thesis.

Isn't the concept of monitoring too broad/underspecified?

Last but not least one might question whether the concept of monitoring may be too broad or underspecified to be testable. On the one hand, it must be admitted that monitoring is a global 'broad' concept, in that it reflects a process of executive control *of* language function. In the hierarchy of cognitive processes, executive control is ordered above language processing. Another way in which monitoring is global is that a single type of control process is employed for the handling of errors at different levels of processing. At the same time and despite this global character, it should be clear from the experimental chapters described above, however, that the Monitoring Theory generates specific predictions with respect to the circumstances under which a monitoring response should be triggered. Importantly, in the different studies the predictions of the monitoring hypothesis differed from those of other reanalysis accounts. This implies that the monitoring hypothesis fulfils an important criterion for a theory namely that of falsifiability.

References

Coulson, S., King, J., & Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, 13, 653-672.

Coulson, S., King, J., & Kutas, M. (1998b). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begeleiter (Ed.), *Evoked Potentials and Behavior* (pp. 13-75). New York: Plenum.

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilisitc word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience,* 8, 1117-1121.

Friederici, A.D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219-1248.

Frisch, A., Kotz, S.A., Von Cramon, D.Y., & Friederici, A.D. (2003). Why the P600 is not just a P300: the role of the basal ganglia. *Clinical Neurophysiology*, 114, 336-340.

Gunter, T.C., Stowe, L.A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.

Hahne, A., & Friederici, A.D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194-205.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159-201.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1-36.

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117-129.

Kutas, M. (1993). In the company of other words: electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533–572.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203–205.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/ SPS-P300 debate. *Language and Cognitive Processes*, 14, 1-14.

Osterhout, L., & Holcomb, P. J. (1995). Event-related brain potentials and language comprehension. In M.D. Rugg & M.G.H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp171-215). New York: Oxford University Press.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147–182.

Van Herten, M., Chwilla, D.J., & Kolk, H.H.J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitve Neuroscience*, 18, 1181-1197.

Van de Meerendonk, N., Vissers C.Th.W.M., Chwilla D.J., & H.H.J. Kolk (submitted). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18.

Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk H.H.J. (2008) Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967-982

Wassenaar, M., & Hagoort, P. (2007). Thematic role assignment in patients with Broca's aphasia: Sentence-picture matching electrified. *Neuropsychologia*, 45(4), 716-740.

Monitoren bij het waarnemen van taal [1]

Het maken van fouten is menselijk. Juist daarom is het belangrijk, dat ons cognitieve systeem zichzelf beschermt tegen de consequenties van fouten. Zo bewaken wij continu de kwaliteit van ons gedrag; hierdoor worden onze prestaties beter (bv., Stuss en Benson, 1986). Voordat een fout gecorrigeerd kan worden, moet deze eerst opgemerkt worden. In het domein van de taalproductie wordt aangenomen, dat er een monitor bestaat om problemen te ontdekken en een proces van correctie op te starten. Monitoren kan hierbij manifest worden in de zogenoemde zelf-reparaties van de spraak. Deze zelf-reparaties voorkomen, dat de spreker verkeerd begrepen wordt door zijn of haar gesprekspartner. We maken ook fouten bij het waarnemen van taal; soms verstaan we iemand verkeerd, of maken we een fout bij het lezen. In dit proefschrift stel ik voor dat er op het gebied van de taalperceptie een monitor werkzaam is vergelijkbaar met die op het gebied van de taalproductie.

Monitoren in taalperceptie

In de literatuur bestaat een groot aantal voorbeelden van spontane mispercepties of ‘ slips of the ear’ (b.v., Cutler & Butterfield, 1992). De zin ‘She’s a must to avoid.’ werd bijvoorbeeld verkeerd gehoord als ‘She’s a muscular boy’. En in een historische anecdote is een Britse officier bekend geworden door de uitspraak ‘Send reinforcements, we’re going to advance’. Deze zin werd verkeerd gehoord als: ‘Send three and four pence, we’re going to a dance.’ Een dergelijke onverwachte interpretatie kan natuurlijk kloppen (ze gingen misschien wel echt naar een feest). Het kan echter ook het resultaat zijn van een perceptuele verwerkingsfout. In mijn dissertatie wordt voorgesteld dat wij perceptuele fouten monitoren om een onderscheid te maken tussen deze twee mogelijkheden. Een dergelijk onderscheid is erg belangrijk omdat het kan bepalen of er zakgeld gestuurd moet worden of extra troepen!

[1] Dit hoofdstuk wordt gepubliceerd als Vissers, C.Th.W.M. (2008). Monitoren bij het waarnemen van taal. *Neuropraxis, 4*.

In het geval van het waarnemen van taal kunnen fouten niet direkt geobserveerd worden zoals dat in gesproken taal mogelijk is (bijvoorbeeld bij een verkeerd uitgesproken woord). Een belangrijke vraag is dus: hoe kunnen taalgebruikers weten dat er een waarnemingsfout heeft plaatsgevonden? Het enige signaal dat ons kan laten weten, dat er misschien iets mis is gegaan bij het verwerken, is een conflict tussen wat je verwacht te lezen of te horen en dat wat je werkelijk leest of hoort. De schending van de verwachting, dat wil zeggen of het conflict tussen wat je verwacht en wat je waarneemt maakt duidelijk dat er sprake is van een mogelijke verwerkingsfout in de taalwaarneming. Volgens de Monitoring Theorie (zie hieronder) ontlokt dit conflict een opnieuw verwerken van de zin om te controleren op de mogelijkheid van een verwerkingsfout.

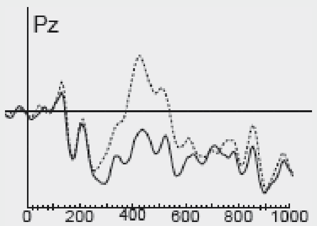
Vensters op het brein: gebeurtenis-gerelateerde potentialen (ERPs)

Omdat de taalgebruiker zelf geen toegang heeft tot de veelal onbewuste cognitieve processen, hebben we voor de experimenten, gepresenteerd in mijn proefschrift, gebruik gemaakt van het EEG (electroencephalogram) van proefpersonen om de cognitieve activiteit tijdens lezen zichtbaar te maken. Het EEG weerspiegelt de algemene toestand van de proefpersoon, zoals bijvoorbeeld wakker zijn of slapen. Wanneer er een grote hoeveelheid neuronen tegelijkertijd actief is tijdens het verwerken van informatie, leidt dat tot gelijkmatige variatie in elektrische activiteit van het brein. Deze elektrische activiteit kan gemeten worden door op de hoofdhuid electroden te plakken. Zogenoemde gebeurtenis-gerelateerde potentialen (ERPs) weerspiegelen kleine veranderingen in de spontane hersenactiviteit van het brein die worden ontlokt door een bepaalde stimulus of gebeurtenis. Normaal gesproken zijn ERP’s niet zichtbaar in het EEG omdat de ERPs veel kleiner zijn in amplitude (5-10 μV) dan de spontane fluctuaties van het EEG (50-100 μV). De ERP kan toch betrouwbaar worden gemeten, als is vastgelegd wanneer een bepaalde stimulus precies wordt aangeboden. Door een groot aantal herhalingen van de stimulus worden de sporen van spontane hersenactiviteit uitgemiddeld. Immers, de elektrische activiteit die gegenereerd wordt door de neuronen die niet betrokken zijn bij het verwerken van de kritische stimulus, is toevallig met betrekking tot het tijdstip van de het stimulusbegin en zal dus uitgemiddeld worden. Zo worden de componenten die aan de stimulus gekoppeld zijn –de ERPs- zichtbaar.

De ERPs, die na middeling zijn vastgesteld bestaan weer uit verschillende ERP componenten die ieder een specifieke verschijningsvorm en verloop hebben en die een bepaald verwerkingsproces in de hersenen weerspiegelen. Zo zijn de N400 en de P600, twee verschillende taalgerelateerde ERP componenten.

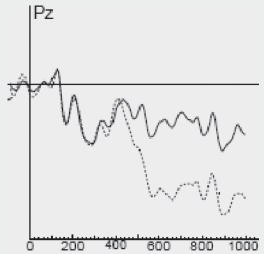
In 1980 lieten Kutas en Hillyard zien, dat woorden die in semantisch opzicht niet passen in een zin, zoals in 'Hij smeerde zijn warme broodje met *sokken*.' een grote negatieve ERP component ontlokken, in vergelijking met semantisch passende woorden, zoals in 'Hij smeerde zijn warme broodje met *boter*.'. Het niet passende woord '*sokken*' roept in vergelijking met het passende woord '*boter*' een N400 effect op. Het venster, de tijdsspanne, van deze negatieve verschuiving ligt rond 400 ms na het zien of horen van het kritische woord (bij deze zin is dat '*sokken*'). Men neemt aan dat de N400 het gemak weerspiegelt waarmee een woord in de context wordt geïntegreerd (bv., Chwilla, Kolk, & Mulder, 2000; Friederici, 1995; Nieuwland & Van Berkum, 2005). Zie figuur 1 voor een voorbeeld van een N400 effect.

Figuur 1 Voorbeeld van een N400 effect. Negatieve amplitudes zijn naar boven uitgezet, positieve amplitudes naar beneden. De zwarte lijn weerspiegelt de reactie op semantische verwachtte woorden, de gestippelde lijn weerspiegelt de reactie op semantisch onverwachtte woorden. De gemiddeldes zijn in de tijd gekoppeld aan het begin van het kritische woord.



Schendingen van de syntactische structuur, zoals in 'Het verwende kind *gooien* het speelgoed op de grond.' ontlokken in vergelijking met de syntactisch correcte zin

Figuur 2 Voorbeeld van een P600 effect. Negatieve amplitudes zijn naar boven uitgezet, positieve amplitudes naar beneden. De zwarte lijn weerspiegelt de reactie op syntactisch correcte woorden, de gestippelde lijn weerspiegelt de reactie op syntactisch incorrecte woorden. De gemiddeldes zijn in de tijd gekoppeld aan het begin van het kritische woord.



een late positieve verschuiving (Hagoort, Brown, & Groothusen, 1993). Het venster van deze positieve verschuiving ligt tussen 500 ms en ten minste 800 ms na het zien of horen van het kritische woord (bij deze zin is dat '*gooien*'). Het P600 effect is ook gevonden na syntactisch ambigue zinnen (bv., Osterhout & Holcomb, 1992) en na syntactisch zeer complexe zinnen (Kaan, Harris, Gibson, & Holcomb, 2000). Er werd aangenomen, dat het P600 effect syntactische heranalyse (Friederici, 1995; Osterhout, Holcomb, & Swinney, 1994), syntactische verwerking (Hagoort et al., 1993) of moeite met syntactische integratie (Kaan et al., 2000) weergeeft. Zie figuur 2 voor een voorbeeld van een P600 effect.

Meer recent echter is het P600 effect ook geobserveerd na semantische schendingen. De zinnen waarin deze effecten werden gemeten voldoen niet aan de hierboven beschreven criteria: ze zijn syntactisch correct, niet ambigu en ook niet bijzonder complex (bv., Kolk, Chwilla, van Herten, & Oor, 2003; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005). Zo vonden Kolk en zijn collega's in 2003 een P600 effect na semantische omkeringen zoals, 'De vos die op de stropers *joeg* sloop door het bos'. Deze semantische omkeringen waren gevormd door het subject en object van de semantisch correcte zinnen om te wisselen. Het vinden van P600 effecten na semantische schendingen in syntactisch niet ambigue zinnen ondergraaft de stelling, dat semantische en syntactische schendingen kwalitatief verschillende ERP patronen oproepen.

De Monitoring Theorie

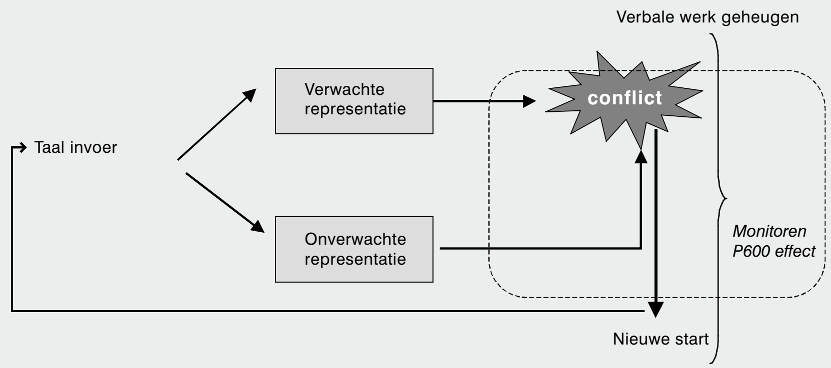
In dit proefschrift werk ik het voorstel uit, dat het P600 effect heranalyse van de zin weerspiegelt met een meer algemene functie dan syntactische herstructurering.

Wij monitoren om te controleren of de oorzaak van een bepaald onverwacht element of van een bepaalde inconsistentie in wat wij lezen misschien ligt in het verkeerd gelezen hebben van bijvoorbeeld een woord of zin. Immers een inconsistentie of onverwachtheid kan twee verschillende bronnen hebben. In sommige gevallen is een onverwachte element ook echt opgetreden, terwijl een inconsistentie in andere gevallen het gevolg kan zijn van een perceptuele fout, zoals hierboven beschreven. Het is net alsof je jezelf afvraagt: 'heb ik dat nou goed gelezen?'

Ik zal nu meer in detail uitleggen hoe volgens de Monitoring Theorie, de lezer omgaat met de consequentie van een mogelijke verwerkingsfout. In Figuur 3 is een schematische representatie weergegeven van de wijze waarop mogelijke fouten in de taalwaarneming ontdekt en gefilterd worden. De gelijktijdige activatie van een verwachte representatie en een onverwachte representatie van dezelfde taaluiting leidt tot een conflict (bijvoorbeeld, bij het lezen van de zin 'de man bijt de hond'. Wij stellen voor dat de lezer in dit geval het goed voor mogelijk houdt, dat hij de zin verkeerd gelezen heeft. Daarom start de lezer met het opnieuw verwerken van de invoer om te controleren of hij de zin goed heeft waargenomen. Het P600 effect weerspiegelt de heranalyse om te controleren op mogelijke verwerkingsfouten. Het verbale werkgeheugen is de arena waar dit soort conflicten wordt opgelost.

Zo stelt de Monitoring Theorie voor, dat het P600 effect na semantische schendingen heranalyse weergeeft, ontlokt door een conflict tussen twee onverenigbare representaties. Bij het lezen van de zin 'De vos die op de jager joeg', stellen de plausibiliteitsheuristiek en het syntactische algoritme verschillende thematische interpretaties voor. De plausibiliteitsheuristiek vormt een interpretatie op basis van de individuele woordbetekenissen; het syntactische algoritme levert een inter-

Figuur 3 Schematische representatie van hoe het taalsysteem omgaat met de consequentie van een mogelijke verwerkingsfout, volgens de Monitoring Theorie.



pretatie op basis van de syntax van de zin. Terwijl in bovenstaande zin de plausibiliteits heuristiek tot de meest plausibele, verwachte interpretatie leidt, namelijk dat 'de jager op de vos joeg', leidt de algorithmische ontleder tot een implausibele, onverwachte interpretatie, namelijk dat 'de vos op de jager joeg.' Het is heel zinnig voor de lezer om de correctheid van zijn of haar analyse te controleren. Immers, zoals eerder gezegd, een inconsistentie kan twee oorzaken hebben. Het kan kloppen, dat een onverwachte gebeurtenis zoals in de zin verwoord, daadwerkelijk heeft plaatsgevonden. Een onverwachtheid kan echter ook veroorzaakt zijn door een perceptuele verwerkingsfout. Het weg filteren van fouten is daarom essentieel voor een goed taalbegrip.

Volgens de Monitoring Theorie weerspiegelt het P600 effect de heranalyse, ontlokt door de activatie van twee onverenigbare representaties. De algemene functie van monitoren is het controleren op mogelijke verwerkingsfouten. Omdat verschillende linguïstische elementen verkeerd kunnen worden waargenomen voorspelden wij, dat het P600 effect heranalyse op een aantal verschillende linguïstische niveaus kan weerspiegelen. Om deze hypothese te toetsen hebben we conflicten ontlokt op drie verschillende niveau's van het linguistische systeem. De resultaten van de experimenten beschreven in dit proefschrift ondersteunen de Monitoring Theorie. De experimenten laten namelijk zien, dat het P600 effect niet alleen heranalyse op het syntactische niveau weerspiegelt, maar ook op drie andere niveau's: namelijk op zinsniveau, woordniveau, en conceptueel niveau.

Monitoren op zinsniveau

In een eerste experiment (Vissers, Chwilla, & Kolk, 2007) hebben we de hierboven beschreven hypothese getest, namelijk dat het P600 effect na semantische omkeringen zoals 'De vos die op de jager *joeg*.' monitoren weerspiegelt ontlokt door een conflict tussen de semantisch plausibele, verwachte thematische interpretatie en de implausibele thematische interpretatie (Kolk et al., 2003). Kolk en zijn collega's stellen voor dat het P600 effect na deze semantische omkeringen de controle op mogelijke verwerkingsfouten op zinsniveau weergeeft.

Deze hypothese wordt onderzocht door hetzelfde stimulusmateriaal aan te bieden als in de studie uit 2003 gebruikt is, maar dan in combinatie met een andere taak instructie. In de eerste studie moesten de proefpersonen aangeven of de zin semantisch acceptabel was of niet (Kolk et al., 2003). In de volgende studie daarentegen werd de proefpersonen verteld, dat de semantische omkeringen met opzet waren gecreëerd (Vissers et al., 2007). Hen werd verteld, dat ze zich niet moesten laten misleiden door hun kennis van wat er normaal gesproken gebeurt in de wereld, maar dat ze extra aandacht moesten schenken aan 'wie wat doet tegen wie.' De rationale achter deze instructie is, dat het de discrepantie tussen de thematische interpretatie voorgesteld door de semantische heuristiek en de syntactische ontleder zou moeten reduceren. Als Kolk en zijn collega's gelijk hebben met hun voorstel, dat het P600 effect na semantische omkeringen is gebaseerd op een controle operatie ontlokt door conflict op zinsniveau, dan zou deze focus-op-syntax instructie moeten leiden tot een afname in fouten en een reductie van het P600 effect.

Onze focus-op-syntax instructie heeft zowel de gedragsdata als de ERP data beïnvloed, zoals verwacht. De proefpersonen in deze studie waren sneller en maakten minder fouten dan de proefpersonen in de studie in 2003. Onze proefpersonen werden dus minder snel misleid door de semantische omkeringen.

De focus-op-syntax instructie leidde ook tot het verdwijnen van het P600 effect op de middellijn en op de linker hemispheer (behalve 1 elektrode positie). Omdat de instructie een onverwachte gebeurtenis ('dat de vos op de jager joeg') in een minder onverwachte gebeurtenis heeft veranderd, is het minder noodzakelijk voor het brein om de aandacht te richten op het implausibele linguïstische element. Dit heeft geleid tot een afname van het P600 effect. De resultaten uit deze studie ondersteunen de hypothese van Kolk en zijn collega's dat het P600 effect na semantische omkeringen is gebaseerd op een controle operatie ontlokt door een conflict tussen twee incompatibele interpretaties op zinsniveau.

Monitoren op woordniveau

In een volgend experiment (Vissers, Chwilla, & Kolk, 2006) hebben we onderzocht of een conflict tussen een zeer verwachte representatie en een onverwachte representatie ook een monitoring response kan uitlokken op woordniveau. Met dit als doel creëerden we zowel een zeer hoge verwachting, (high-cloze context; 'In die bibliotheek lenen scholieren *boeken* om mee naar huis te nemen.') als een lage verwachting (low-cloze context; 'De kussens zijn opgevuld met *boeken* waardoor ze hard aanvoelen.') voor een bepaald lexicaal item. Het kritische lexicale item was ofwel correct gespeld ofwel een pseudohomofoon. Pseudohomofonen zijn afgeleid van het correcte woord en fonetisch gelijk aan dit woord; echter, ze zijn verkeerd geschreven (bv., *boekun*).

We voorspelden dat de pseudohomofoon/high-cloze zinnen zouden leiden tot een conflict op woordniveau, tussen de neiging om de pseudohomofoon te accepteren en de neiging om deze af te wijzen. Er werd verondersteld, dat de neiging om de pseudohomofoon te accepteren zeer sterk is. Ten eerste, omdat de pseudohomofoon correspondeerde met een woord dat semantisch zeer verwacht is. En ten tweede, omdat de fonologische vorm van de pseudohomofoon deze semantische verwachting nog eens bevestigde. Aan de andere kant is de neiging om de pseudohomofoon af te wijzen ook sterk omdat het een verkeerd gespeld woord is. De verwachting was dat het resulterende conflict het brein in staat van besluiteloosheid brengt en monitoren ontlokt, en dat dit wordt weerspiegeld in een P600 effect. In de low-cloze zinnen waren de lexicale items waar de pseudohomofonen van zijn afgeleid niet verwacht.

Ze werden daarom verondersteld geen of een minder groot conflict te ontlokken tussen een verwacht en een gepresenteerd lexicaal item. Er werd voor de low-cloze zinnen dus ook geen of een minder groot P600 effect verwacht.

De ERP data bevestigden onze voorspelling dat alleen pseudohomofonen in een high-cloze context een P600 effect ontlokken. Omdat de woorden in ons experiment waar de pseudohomofonen van zijn afgeleid zeer verwacht zijn, worden de pseudohomofonen in eerste instantie gemakkelijk geïntegreerd in de representatie van de context op een hoger betekenisniveau. Echter, monitoren wordt ontlokt wanneer de lezer de verkeerde spelling ontdekt. Het conflict tussen het verwachtte woord en het verkeerd gespelde woord signaleert immers de aanwezigheid van een mogelijke verwerkingsfout. Het monitoren werd ook hier weerspiegeld in het P600 effect. Omdat de pseudohomofonen en woorden niet verwacht waren in de low-cloze zinnen, werden ze niet gekwalificeerd als verwerkingsfouten en verwarden ze de lezer ook niet. Daarom werd in de low-cloze zinnen geen monitoren of P600 ontlokt.

Deze resultaten ondersteunen de notie dat er in aanvulling op monitoren in taalperceptie op zinsniveau, ook monitoren op woordniveau geschiedt. Monitoren vindt plaats wanneer een conflict tussen een sterke neiging om een woord te accepteren en de neiging om het woord af te wijzen, het cognitieve systeem in staat van besluiteloosheid brengt. Het brein controleert de invoer om het conflict tussen twee incompatibele representaties op te lossen.

Monitoren op conceptueel niveau

In een derde experiment (Vissers, Chwilla, Van de Meerendonk, & Kolk, 2008) hebben wij op een geheel andere wijze een hoge verwachting opgeroepen, namelijk door het aanbieden van plaatjes. Zo hebben we onderzocht of een conflict tussen een zeer verwachte representatie en een onverwachte representatie ook monitoren kan ontlokken op conceptueel niveau na een mismatch tussen een plaatje en een zin. Dit werd getest door de verwachting te manipuleren en de waarheidswaarde tussen een plaatje en een zin te manipuleren. We hebben de proefpersonen plaatjes van ruimtelijke figuren gepresenteerd. Deze plaatjes werden weggehaald en gevolgd door een zin die afhankelijk van het gebruikte voorzetsel een correcte

of een incorrecte beschrijving van het plaatje gaf. In de mismatch condities (de verkeerde combinatie conditie) beschrijft de zin het plaatje niet correct. We hebben twee verschillende soorten mismatches gebruikt: namelijk, intra-dimensionele mismatches (bv., □△ - 'De driehoek staat *voor* het vierkant.') en extra-dimensionele mismatches (bv., □△ - 'De driehoek staat *boven* het vierkant.'). We voorspelden, dat de mismatches tussen een plaatje en een zin een conflict oproepen tussen de verwachte conceptuele representatie gebaseerd op het plaatje en de onverwachte conceptuele representatie afgeleid van de zin. Ook dit conflict zou monitoren weerspiegeld in een P600 effect moeten ontlokken

Zoals voorspeld, werd een robuust P600 effect ontlokt na zowel de intra-dimensionele mismatches en de extra-dimensionele mismatches. Het P600 effect in beide condities was breed verdeeld over de schedel. Het P600 effect in beide condities laat een centroparietaal maximum zien, laatstgenoemde verdeling is ook kenmerkend voor P600 effecten in reactie op allerlei syntactische afwijkingen. We stellen dat de onderliggende uitlokker voor het P600 effect weer de aanwezigheid is van een conflict. In dit geval, een conflict tussen de conceptuele representatie gebaseerd op het plaatje en de conceptuele representatie gebaseerd op de zin.

Conclusies

Zoals hierboven beschreven leert de Monitoring Theorie ons niet alleen iets over de manier waarop perceptuele fouten worden ontdekt en gerepareerd, het werpt ook nieuw licht op de functionele betekenis van het P600 effect. Voorstanders van de huidige theorieën houden vol, dat het P600 effect syntactische herverwerking weerspiegelt, met als doel het verwijderen van een syntactisch obstakel zoals een grammaticale schending of een syntactische ambiguiteit. Wij zijn ook van mening dat de P600 het opnieuw analyseren weerspiegelt. Echter, wij stellen dat het doel van deze heranalyse niet puur syntactisch is, maar meer algemeen, namelijk om de kwaliteit van de waarneming te evalueren. In de hierboven beschreven studies hebben we laten zien dat het P600 effect ook heranalyse kan weerspiegelen op zinsniveau, woordniveau, en conceptueel niveau. De functionele interpretatie van het P600 effect moet derhalve uitgebreid worden van puur syntactische heranalyse naar meer algemene heranalyse of monitoren.

Monitoren is een centraal aspect van cognitieve of executieve controle. Executieve controle is steeds nodig als het taalsysteem in een staat van onzekerheid is; dit is bijvoorbeeld het geval als meerdere zinsrepresentaties tegelijkertijd actief zijn. Men kan executieve controle zien als extra aandacht die ingezet wordt om het conflict op te lossen. Die extra aandacht kan een bepaalde keuze vergemakkelijken wanneer er bij het lezen conflicterende alternatieven zijn, of het kan ingezet worden om het geheugenspoor van een woord of plaatje terug te halen om te controlen of het wel goed verwerkt is. Het doel van deze executieve controle is de optimalisatie van taalverwerking.

Referenties

Chwilla, D.J., Kolk, H.H.J., & Mulder, G. (2000). Mediated priming in the lexica decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, 42, 314-341.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.

Friederici, A.D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, 50, 259-281.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.

Hoeks, J.C.J., Stowe, L.A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59-73.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159-201.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85,1-36.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203–205.

Nieuwland, M.S., & Van Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24, 691-701.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(4), 786-803.

Stuss, D.T., & Benson, D.F. (1986). *The frontal lobes*. New York: Raven Press.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150-163.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18.

Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk H.J. (2008) Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967-982

Dankwoord

Dit boek is bijna uit, maar niet voordat ik eenieder heb bedankt voor zijn of haar unieke bijdrage.

De eerste woorden van dank zijn gericht aan mijn ouders. Lieve mama, samen met papa schonk je de kleurrijke woorden en onvoorwaardelijke liefde die mij hebben doen groeien tot wie ik nu ben. Jouw zuivere intuïtie was voor mij steeds een vast baken. Met jouw warmhartige zorgzaamheid en aanstekelijke humor maak jij een fijn thuis voor ons allemaal. Lieve papa, al heel vroeg wist jij mijn interesse voor filosofie en psychologie te prikkelen met jouw sprekende vertalingen naar mijn jonge belevingswereld. Met het jou kenmerkende kritische inzicht en het vermogen nieuwe perspectieven te scheppen, creëerde jij voor mij een veilige en inspirerende denkwereld waar ik nog altijd vol vertrouwen op terug kan vallen. Met veel liefde en dank draag ik dit proefschrift aan jullie op.

Voorts wil ik mijn co-promotor danken. Lieve Dorothee, zonder jouw voortreffelijke expertise als wetenschapper, jouw aanmoedigende woorden en verbindende toewijding, was het project niet in deze vorm voltooid. De manier waarop jij over mijn wetenschappelijke maar ook over mijn persoonlijke groei waakte was geweldig inspirerend en stimulerend. Onze hechte samenwerking en oprechte vriendschap zijn mij bijzonder dierbaar. Grote dank ben ik tevens verschuldigd aan mijn promotor. Beste Herman, jij hebt mijn wetenschappelijke verkenningen met vertrouwen en vruchtbaar enthousiasme omgeven. Al tijdens mijn eerste colleges Geschiedenis van de Psychologie raakte ik geboeid door jouw geanimeerde beschrijvingen van wetenschap en psychologie. Het is mij nooit gelukt om aan jouw kritische beschouwingen te ontkomen; des te beter, want mede daardoor bleef ik aan het denkwerk. Ik heb met volle teugen genoten van de logische argumentaties die samen met jou en Dorothee tijdens de levendige besprekingen werden opgebouwd. Op deze wijze heeft het grondthema van dit project zich verdiept.

I am indebted to the members of the manuscript committee, Laurie Stowe, Lee Osterhout and Herbert Schriefers for their careful reading of the manuscript. I would especially like to thank Lee Osterhout for the generous hospitality he extended to Maarten and I at the Cognitive Neuroscience Lab in Seattle.

Ik mag me gelukkig prijzen dat er fijne collega's om mij heen stonden die in het zelfde schuitje voeren. Lieve Nan en Marieke, het was waardevol en vermakelijk om met jullie te stoeien over de experimenten en te reflecteren over de Monitoring Theorie. En Nan, bedankt voor de plezierige samenwerking en jouw bijdrage aan het plaatjes experiment. Lieve Kathleen, jij bracht een opgewekte stemming en liet gelijk mijn Engels zoveel beter klinken. Jouw inspanningen zijn de inhoud van dit werk zonder twijfel ten goede gekomen. Let's keep in touch! Lieve Tessa, als NICI collega ging jij mij voor naar het Radboud; dankjewel voor jouw hartelijke belangstelling bij mijn overstap en de verkwikkende lunchafspraken. Beste Martijn, Edita, Arno, Mark, Ellen, Kristin, Stan, Janneke en andere junior-onderzoeker collega's; veel dank voor de ontspannen(de) gesprekjes tussendoor. Lieve Evelien, met jou deelde ik naast heerlijke cappuccino's, de kleine en grote gedachten van de dag. We vertrouwden elkaar veel toe, en als je kon dan sprong je bij. De open gesprekken met jou over het zweet en de tranen die wetenschap soms kan voortbrengen blijven voor mij van grote betekenis; mede dankzij jouw gevoel voor humor en verbale talent komen we vroeg of laat toch tot relativering (de wereld is geen toverbal). Lieve Inge, beste kamergenote; wat fijn dat jij Sammie's vrolijke doch harige aanwezigheid kon waarderen en dat Roy en jij zulke bouwexperts zijn, want hulp komt soms uit onverwachte hoek. Dankjewel voor de vele lachbuien, voor de knappe praatjes over het gewone leven en voor het nakijken van de allerlaatste versie van mijn proefschrift! Lieve Kim, met jou heb ik warme, eerlijke verbondenheid ervaren, als collega èn als vriendin. Wat hebben we tijdens die sfeervolle avondjes genoten van jullie lieve gastvrijheid en Dons uitzonderlijke kookkunsten. Amsterdam, wij komen eraan!

Bij het opzetten, uitvoeren en analyseren van de EEG experimenten heb ik hulp gekregen van mensen die bereid waren hun deskundigheid aan te wenden om onze 'experimenteer' ideeën in een goed werkende EEG opstelling om te zetten. Ik wil hen hier dan ook dankbaar vermelden: Hubert, Jos, Pascal, Gerard, Andre en Lee.

Mijn dank gaat ook uit naar Beppie van Dijk, Thea Holla en Yvonne Schouten (pruimen in Slibovitz zijn echt heerlijk!) voor de administratieve ondersteuning. Naast de serieuze regelzaken zorgden jullie op langzame momenten voor een frisse wind door de gang.

Graag wil ik Angelique Hendriks noemen. Jij bracht jouw expertise in om mij een rechtstreekse introductie te geven in het spannende oogbewegingonderzoek. Dankjewel voor jouw intensieve betrokkenheid en eindeloze geduld.

De afgelopen jaren heb ik veel tijd en inspanning besteed aan het geven van onderwijs aan psychologiestudenten. In het bijzonder wil ik Ben Hofstede danken voor de geboden kans om mijn junioren tijd op deze manier te verrijken en voor het in mij gestelde vertrouwen.

Wendy Bergervoet mag zeker niet vergeten worden; dankjewel voor de creatieve en geduldige manier waarop jij het portret door Picasso hebt geïntegreerd in de omslag. Dank ook aan Harald Pieper, voor het lay outen van mijn boekje. Ik ben erg blij met het resultaat.

Mijn lieve vriendinnen ‘buiten de wetenschap’ wil ik bedanken voor hun trouwe vriendschap ook toen ik veel te druk was. Lieve Maartje, met intens plezier en weemoed denk ik terug aan jouw wekelijkse maandag-nacht-verblijf op gang 200 en aan onze fanatieke pogingen tot watersport. Ik waardeer jouw heldere blik en attentheid enorm. Lieve Mirjam, soms was je even uit zicht, maar gelukkig blijven onze vriendschap en onderhoudende uitstapjes in Arnhem voortduren. Lieve Jennifer, onze vriendschap stamt al vanaf de eerste klassiek balletlessen en groeide uit op het gymnasium (quatre-mains!). Ik ben blij dat wij elkaar nog steeds weten te vinden.

Dan wil ik mijn grote zussen, broer en zwagers danken. Lieve Honorée en Arthur, ook al bleef jullie kinderschare groeien, we hebben steeds onmisbare bijstand in woord en daad ervaren van jullie. Het is heerlijk om bij mijn eerste stappen in de klinische wereld het luisterend oor van een expert te mogen hebben. Lieve Antoinette en Marc, met jullie beleefden we naast de grenzen van super spierkracht (I love beton?) ook oergezellige momenten die mij de proefschrift spanning compleet deden vergeten. Zet Lupa maar vast klaar, want als dit werk af is, wil ik samen paardrijden! Lieve Wynand, mijn grote broer; wat hebben wij vroeger veel gesteggeld én plezier beleefd tijdens ons cello/piano spel. Jij bent voor mij een inspirator en dat jij nog altijd een oogje in het zeil houdt geeft een warm gevoel van geborgenheid. Jij en ik delen de passie om tot het wezen van een fenomeen door te dringen, daarom is het geweldig dat jij de rol van paranimf vervult.

De allerlaatste woorden zijn voor jou, liefste Maarten. Ik ben je onuitsprekelijk dankbaar voor de liefdevolle en groothartige manier waarop je mij ter zijde staat. Het is reuze dat jij als trouwste reisgenoot mijn andere paranimf bent. Samen met jou genieten, dromen en bouwen aan wat er voor ons in het verschiet ligt maakt voor mij dat *het zijn* klopt.

Constance, Huisseling, 4 juni 2008

List of Publications

Publications in international peer-reviewed journals

Vissers, C.Th.W.M. (2008). Monitoren bij het waarnemen van taal. *Neuropraxis, 4*.

Van de Meerendonk, N., Vissers, C. Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (submitted). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns.

Vissers, C.Th.W.M., Kolk, H.H.J., Van de Meerendonk, N., & Chwilla, D.J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967-982.

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8-18

Chwilla, D.J., Kolk, H.H.J., & Vissers, C.Th.W.M. (2007). Immediate integration of novel meanings: N400 support for an embodied view of language comprehension. *Brain Research*, 1183, 109-123

Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J. (2006). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *Brain Research* 1106, 150-163.

Published Abstracts

Van de Meerendonk, N., Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J (2008). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience,* 124. Presentation at the Annual Meeting of Cognitive Neuroscience Society, San Francisco, USA.

Vissers, C.Th.W.M., Chwilla, D.J., Van de Meerendonk, N., & Kolk, H.H.J (2007) Reanalysis in Language Perception: Evidence from ERPs in a Sentence-Picture Matching Task. *Journal of Cognitive Neuroscience*, 33. Presentation at the Annual Meeting of Cognitive Neuroscience Society, New York, USA.

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2005). A Semantic Bias in Sentence Perception. *Journal of Cognitive Neuroscience*, 35. Presentation at the Annual Meeting of Cognitive Neuroscience Society, New York, USA.

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2003). A monitoring process in language perception: Evidence from Event Related Potentials. *Journal of Cognitive Neuroscience*, 93. Presentation at the Annual Meeting of Cognitive Neuroscience Society, San Fransisco, USA.

Chwilla, D.J., Kolk H.H.J., & Vissers C.Th.W.M. (2003). Tracking the time-course of the creation of new meanings. *Journal of Cognitive Neuroscience*, 175. Presentation at the Annual Meeting of Cognitive Neuroscience Society, San Fransisco, USA.

Presentations

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2007). Monitoring in language perception: evidence from ERPs. University of Washington, *Seattle*

Van de Meerendonk, N., Vissers, C.Th.W.M., Chwilla, D.J., & Kolk, H.H.J.(2007) Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *11th NVP winter conference, Egmond aan Zee.*

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2007). Reanalysis in language perception: evidence from ERPs. FCDC Lecture, *Nijmegen*

Vissers, C.Th.W.M., Kolk, H.H.J, Hendriks A.W.C.J., & Chwilla, D.J. (2006). Monitoring after a conflict: Evidence from Event Related Potentials and Eye tracking. *Amlap, Nijmegen*

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2004). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *NVP winterconferentie, Egmond aan Zee*

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2004). A monitoring process in language perception: Evidence from Event Related Potentials. NBLC, *Doorwerth*

Vissers, C.Th.W.M., Kolk, H.H.J, & Chwilla, D.J. (2003). A semantic bias in sentence perception. Is it subject to control? *NVP winterconferentie, Egmond aan Zee*

Curriculum Vitae

Constance Vissers is op zaterdag 3 november 1979 om 10 over half 12 s'avonds, geboren te Bergen op Zoom. Vanaf 1992 bezocht ze het Juvenaat H. Hart RK Gymnasium te Bergen op Zoom waar ze in 1998 haar diploma behaalde. Hierna is ze in de voetsporen van haar vader en oudste zus getreden en begonnen aan de studie psychologie te Nijmegen. Vanuit interesse voor onderzoek en onderwijs werkte ze vanaf haar tweede studiejaar onafgebroken als student-assistent op de afdeling mathematische psychologie en op de afdeling neuro- en revalidatie psychologie, waar zij zowel op het terrein van onderwijs als op het terrein van onderzoek taken had. In september 2003 studeerde ze cum laude af in de neuro- en revalidatie psychologie. Aansluitend is ze als junior onderzoeker begonnen bij het Nijmegen Instituut voor Cognitie en Informatie (NICI) onder begeleiding van prof. dr. H. Kolk en dr. D. Chwilla. Gedurende deze tijd heeft zij ook steeds onderwijs gegeven aan psychologiestudenten. Het resultaat van haar promotieonderzoek staat opgetekend in dit proefschrift. In 2006 is haar het Frye-stipendium toegekend. Het stipendium heeft zij aangewend om in de zomer van 2007 een bezoek te brengen aan het Cognitive Neuroscience of Language Lab van de University of Washington, Seattle. Vanaf oktober 2007 werkt Constance als psycholoog en postdoc onderzoeker bij het Universitair Medisch Centrum St Radboud op de afdeling Psychiatrie. Vanaf januari 2008 is zij daar tevens in opleiding tot gezondheidszorg psycholoog.